

18/25 December 1980

Britain, France and nuclear weapons

Between now and 20 January, when Mr Ronald Reagan will be inaugurated as president of the United States, much important business elsewhere will be conducted in a vacuum. So much is clear from last week's meeting of the NATO Council in Brussels. Many of those present seem to have been as interested to know whether their old colleague, General Douglas Haig, will finish up as United States Secretary of State as in the prospect that threats of economic sanctions against the Soviet Union would deter an invasion of Poland. Yet vacua have their benefits. As boxers tend to free-associate in the nervous minutes before the beginning of a fight, so politicians and others in the West have been talking openly about some of the issues on which American influence will again be dominant in January — defence, arms control and the like. In Britain, this unaccustomed free speech has been mixed with the wild talk inseparable from the accession of Mr Michael Foot as leader of the British Labour Party, for Mr Foot used to be a staunch supporter of unilateral nuclear disarmament for Britain, and has not yet renounced his former views. The Campaign for Nuclear Disarmament can boast of more recruits in the past few months than in the past several years.

Economic issues apart, the most obvious uncertainty is Mr Reagan's position on arms control, and on his plan to renegotiate the Salt II agreement. During the election campaign, his position shifted steadily, from outright repudiation to renegotiation. Soon after 20 January, it should be possible to tell where he now stands, but the signs are encouraging. The Soviet Union has not slammed the door on further talks about the agreement that President Carter shrank from fighting through the US Senate. His successor must by now know the strength of opinion in Europe, especially in Germany, about the need for agreement on short-range as well as strategic nuclear weapons. Anxiety about Poland has only sharpened suspicions that the stability engendered by mutual nuclear deterrence has its drawbacks. With nothing worth mentioning happening at the Madrid conference on the Helsinki accords, Mr Reagan is likely to have calculated that arms control is not, as he once supposed, the wishy-washy plaything of liberals but a considerable political opportunity. That, at least, must be the hope.

Will governments elsewhere be capable of similar (and still unproven) flexibility? The need is most urgent in Britain, as the defence debate in the House of Lords on 3 December showed. Since the invention of nuclear weapons, there have been two British positions on nuclear defence — the belief (shared by all governments in office) that an independent nuclear deterrent is a necessary part of British policy. This is the belief which led the Callaghan government to fit the Polaris submarine missiles with new warheads (at a cost estimated at £1,000 million) and the present government to order a fleet of six Trident nuclear submarines (at a cost of £5,000 million spread over the next ten years). At the other extreme is the view that British bombs are not a strength but a weakness, and should be dismantled. The unilateral position is not homogenous — some ask merely for the abandonment of British nuclear weapons, others would refuse to provide bases for US nuclear weapons or even withdraw from NATO. Nor are the unilateralists entirely wrong-headed in their views. Unilateral nuclear disarmament would bring certain benefits. The snag, which hitherto has far outweighed the advantages, is that there would be a profound change in the character of British political relationships across the Atlantic and across the Channel — and no assurance of British immunity from the consequences of foreign wars.

The novel development is that, in Britain, there is a growing body of opinion sandwiched between the two conventional extremes. In the House of Lords debate, people such as Lord Zuckerman (once Chief Scientist at the Ministry of Defence) and Lord Chalfont (Mr Harold Wilson's Minister for Disarmament) poured cold water on the decision to replace the Polaris submarines with Trident. "If we can afford it, so be it", said Lord Zuckerman, who knows as well as the next British taxpayer that the government's survival will soon depend on its success in cutting public expenditure. The case against the Trident replacement unites an impressive range of interests. The military (represented, for example, by Lord Carver, a distinguished field marshal) would prefer to see the Trident money spent on conventional forces. Strategists (like Chalfont) argue that weapons other than Trident missile submarines would be cheaper and no less effective. None of this implies support for unilateral disarmament — the denial of British bases to US nuclear warheads (in aircraft now, in cruise missiles sometime after 1983), for example. What has happened is merely that entirely sensible opinion has begun to question the most conspicuous cornerstone of British defence policy in the period since 1945.

The British government has been slow to see the way the wind is blowing, or to recognize the advantages of following Mr Reagan's apparent willingness to make the best out of accommodations with his critics. The most immediate embarrassment of the Trident programme is its cost, probably grossly underestimated at £5,000 million. Can it make sense that a government whose chief failure since its election eighteen months ago has been its failure to cut public expenditure should now take on a capital commitment on such a scale, more than half the cost of the US Apollo programme? The decision might be more easily justified if (as some Apollo-builders held in the 1960s) building the Trident submarines would bring uncovenanted technical innovations, but the objective is largely to replicate the nuclear submarines already being built in the United States. It is no wonder that there has been a revival, in the past few weeks, of the old dream of an Anglo-French cruise missile system. That would be cheaper and politically powerful.

These are narrow arguments. The broader question of whether countries such as Britain — France is the only other example — should hang on to independent nuclear forces is more divisive. The conventional British statement of the case is that nuclear weapons are a contribution to collective defence and an assurance of influence in international negotiations, about disarmament as well as making war. The French view, more iconoclastic, is that French security depends in the last resort on French arms. Both arguments are stiffened by chauvinism. With time, each is weakened. Each of them is less convincing as strategic technology becomes more sophisticated. Yet in Western Europe there is a crying need that the European Community should soon acquire what it should always have had — a foreign policy and a defence policy to go with it. Is that where Britain and France may pool their nuclear ambitions after 20 January?

Nature publication date

As in previous years, this issue of *Nature* (dated 18/25 December) is the last to be published in 1980. Because there are 53 Thursdays in 1981, the next issue, published on 8 January, will be dated 1/8 January 1981.



US research budget now in better shape

Dying Congress restores earlier cut-backs

Washington

It may not last, but in contrast with previous years, when President Carter's requests for additional funding for science were trimmed back by Congress, the final 1981 research budget looks in even better shape than when it was presented in January 1980.

The Administration revised its initial request in March, reducing a proposed growth in spending on research and development from 11 to 8 per cent. But while Congress has agreed to cuts in other areas of social spending, research and development has been relatively protected.

Last week, for example, when Congress finally agreed to an appropriations bill covering the budgets for 23 assorted federal agencies, only three — the National Aeronautics and Space Administration (NASA), the National Science Foundation (NSF), and the American Battle Monuments Division — received more money than they had asked for.

As in previous years, legislators have been kind to the budget of the National Institutes of Health (NIH). Although the final figure is yet to be agreed, the House has approved an NIH budget of \$3,616 million, and last week the Senate gave its approval to a \$3,686 million budget. Whichever figure is finally decided upon, it will be higher than the \$3,581 million initially requested, and will result in a 5.5 to 7.5 per cent increase in NIH's budget over last year.

In the Department of Energy, threats to the basic research budget — which includes funding for high-energy physics and nuclear physics — were successfully fought off during the summer. The final sum for items under the responsibility of the department's Office of Energy Research will be \$1,145 million, only slightly less than the \$1,203 million requested, with the major reduction being the postponement of improvements to facilities at the national research laboratories.

Only in the case of the Department of Defense, for which the Administration had originally requested a hefty 17 per cent increase in spending on basic research, is the final figure likely to be substantially lower. Congress has approved an appropriations bill which will cut this increase to 12.5 per cent, reducing the planned real increase from 8 to 3.5 per cent. The final figure may be even less, since Congress has also asked the Pentagon to impose further unspecified cuts totalling 4

per cent across the whole of its research and development budget. In the case of NASA and NSF as well, the final figures may be reduced by a requirement that the Office of Management and Budget distributes a further 2 per cent cut across all the agencies covered in the appropriations bill.

Congress has also forced several significant changes in the content of the Administrations proposals for federally-sponsored research programmes.

At NSF, for example, sustained pressure from Senator Edward Kennedy has led to the agreement to introduce a new programme to boost women and minority scientists. Mr Kennedy had added his own "Women in Science" bill to the Senate version of the NSF appropriations bill; and the House finally agreed to a \$30 million programme which will include more than 100 grants for women scientists extra to those which would be otherwise awarded. The NIH budget, on which a final total has yet to be settled, includes an extra \$35 million for training grants which the Administration had suggested be cut back to enable the number of new and

competing research grant awards to be kept constant.

A large part of the increase in NASA's budget will go to cover the additional costs of the much-delayed space shuttle. However, Congress is also becoming increasingly concerned about cost and schedule over-runs in virtually all of the space science programmes.

As a result, Congress has directed NASA to ensure that it receives approval from the National Academy of Sciences before making any substantial programme changes. This move, opposed by several NASA administrators, is supported by Congress as a way of ensuring an adequate response to any future technical difficulties.

Despite the set-backs being experienced by the various space science programmes, almost all have escaped the congressional budget process relatively unscathed, including the proposed new gamma-ray observatory. However, there were warnings that some projects could still be halted through budget rescissions early next year when the new administration takes over.

Oil, shale or mirage in West Siberia?

The strange story of the supposed Bazhenov oil field in western Siberia, with its reserves of over 600,000 million tonnes of oil, reached the world's newspapers two weeks ago at a particularly sensitive moment. The threat to world oil supplies posed by the Gulf War, and Western contingency plans for a further technological embargo on the Soviet Union should Soviet troops be ordered into Poland, and the impending OPEC meeting in Bali made the announcement of the new field, with reserves surpassing world proved oil reserves, a major consideration for politicians and oil magnates alike.

The report on the Bazhenov field came from Petrostudies, a two-man Swedish team which monitors all open Soviet publications on the oil industry. The consensus of expert opinion, ranging from the CIA to the Soviet oil industry, is that its report is an exaggeration. One French expert went as far as saying that it would amount to a "geological miracle". Petrostudies, however, remains adamant about the size of the find, but is now prepared to concede that a considerable proportion of the oil may not be recoverable.

Challenged with a Soviet statement that the deposit was not oil but shale, Petrostudies director Manlyo Jermol replied that the Bazhenov shale was scientifically important because it contained liquid crude. Until now, he said, it had been believed that such oil was to be found only in porous rocks. He cited the Soviet

geologist, Ivan Nesterov, director of the West Siberian Research Centre for Oil Prospecting who, in 1978, had announced that Soviet research had refuted the old text-book assertion that shales cannot accumulate free oil.

A few days before the Petrostudies announcement, Nesterov was interviewed on Moscow radio in connection with the publication of the guidelines for the new Five Year Plan. On this occasion he was vague about the West Siberian potential. The fields now in production yield more than 300 million tonnes of oil and gas condensate per year, he said — amounting to about half the total Soviet output. But since two thirds of all the fields discovered in the area have not yet been touched, and only one fifth of the oil-bearing region of West Siberia has been explored it is difficult to make reliable estimates of the reserves. Extraction costs, he pointed out, are soaring, the costs of erecting a rig have risen 30 per cent in the last five years, and it has proved impossible to sustain the original growth rate. While maintaining that Western talk about Soviet oil running out in the near future is "highly premature", Nesterov's talk was couched in cautious terms. During the next five years, he said, the Soviet Union will be able to meet its own requirements and its commitments to its trading partners. For the longer term, he gave no hint of any bonanza — shale or otherwise, of the type suggested by Petrostudies.

Vera Rich

The one potential casualty, the National Oceanic Satellite System (NOSS) — the proposed successor to the SEASAT satellite — will receive an extra \$6.4 million through the research budget of the Department of Commerce's National Oceanic and Atmospheric Administration (NOAA). NOAA has also received an additional \$1 million for manned undersea facilities, but had \$2 million for acid rain research cut from its budget.

In biomedical research, Congress has rejected proposals from both the House and the Senate for greater supervision of NIH. A bill continuing authorization for the National Cancer Institute and the National Heart, Lung and Blood Institute was passed on 5 December without either the provision for a Presidential Commission on biomedical research priorities, proposed by Senator Edward Kennedy, or controversial new authorizing legislation for the remaining nine research institutes which had been proposed by Representative Henry Waxman.

Cuts in the Defense Department's proposed research programme could set back the growth of ties between the military and universities that have been developing over the past four years, according to Pentagon officials, who warned that in particular proposed new programmes for the Department of the Army may have to be cut back.

Nor will the Defense Department be receiving the hoped-for funds to construct a new production facility for binary chemical weapons. This was deleted by members of the Senate from next year's appropriations bill, on the grounds that if the United States is to resume chemical weapons production, this should be a presidential not a congressional decision.

David Dickson

Nuclear safety

European hazards

Brussels

European nuclear power stations are not as vulnerable to operational errors as the one involved in the accident at Three Mile Island, but there is plenty of room for improvement in safety measures in nuclear power stations in the European Economic Community. This conclusion comes from a report submitted to the European Commission's Interdepartmental Coordinating Committee on Nuclear Safety (CCNS). Many of the report's proposals are disputed by the committee in an accompanying reply.

The report comes from a four-man group set up by the European Commission after the Three Mile Island incident. The members of the group were H. Dunster, Deputy Director General of the UK Health and Safety Executive, Professor Latzko of the Technische Hogeschool Delft, Professor Smidt of the Institut für Reaktor Technik der Universität Karlsruhe, and Mr

S. Villani, Director General of the community's Joint Research Centre.

Conditions similar to those which led to the Three Mile Island accident have occurred on several occasions in Europe — the report criticizes the fact that these went almost unnoticed, and calls for a data bank recording all abnormal events in European nuclear power stations.

The siting of nuclear power stations is becoming increasingly politically sensitive in Europe. The report stresses that siting has only a limited part to play in protecting the population, but that there should be a consistent approach among member states to siting nuclear plants, especially in areas close to national borders.

The group also considers that more should be done to minimize the effects of any accidents that may occur: the Commission should study the emergency procedures operating in the various member states, there should be a review of the emergency plans made by the power station operating organizations before they are licenced to operate, and more attention should be given to ways of keeping the public informed in the event of an accident.

The European Commission is criticized for not responding rapidly enough to changing nuclear research needs, for setting up cumbersome committees, and for providing those committees with inadequate technical and administrative services.

According to the Interdepartmental Coordinating Committee, many of the recommendations in the report actually tie in with actions already taken by the Commission, or actions already under consideration. But the rather half-hearted attitude towards nuclear power expressed in the report is not welcomed by the committee. The report says: "no amount of care will totally eliminate the risks of this (nuclear), or any other sort of energy . . . (but) . . . we are finally led to the belief that nuclear sources should continue to play a significant part in the supply of Europe's energy."

Jasper Becker

Fast reactors

Low morale

Staff morale may be as much a threat to the British fast reactor programme as the prospect of a public inquiry on the project. The latest sign of this is the resignation of Mr Jack Moore, coordinator of the fast reactor programme at the UK Atomic Energy Authority (UKAEA), at the end of the year. Mr Moore, who is 57, is leaving to take up a post with Motor Columbus engineering consultants in Switzerland. He said earlier this week that at UKAEA he was unlikely to see his work of the past seven years come to fruition before his retirement.

Mr Moore's resignation highlights two potential problems for the staffing policy of the fast reactor team. Although no other

senior staff are reported to be leaving, further delay in a commitment to build a fast reactor may prompt others to go. The second problem is that of the age structure of the design team. Although the UKAEA has been expanding the team by bringing in young people, previous recruitment policies have left a noticeable dearth of people in their forties. When the senior staff retire or leave, their posts will have to be filled by much younger people.

A government statement on the commercial demonstration fast reactor has been expected since the summer. Sir John Hill, chairman of the authority until the end of the year, submitted a proposal to the government a year ago. Mr Moore expects that the government will respond shortly but that it will not make a final decision at this stage. One complication is that the government has not yet responded to the French proposal that Britain should buy into commercial exploitation of Super-Phénix.

Meanwhile, the Nuclear Power Company is nearing the completion of a detailed design study, which Mr Moore claims will be superior to current French and Russian designs. The UKAEA would like to submit it to the Central Electricity Generating Board and potential reactor manufacturers so that a site could be chosen and a total project proposal put to the government.

Judy Redfearn

Satellite communications

Free for all ahead?

Washington

While the future of US remote-sensing satellites remains entangled in controversy (see *Nature* 14 August), the use of telecommunications satellites is poised for a dramatic expansion.

The Federal Communications Commission (FCC) in Washington has given permission for the launching of 20 new domestic communications satellites, which are likely to increase the capacity of the present system by a factor of four by the mid-1980s.

At least one newspaper company is discussing plans for a nationwide system of locally-produced newspapers linked by satellite, while the Communications Satellite Corporation (Comsat), which owns three of the nine communications satellites at present in orbit, has proposed starting a four-satellite system beaming television programmes directly to private homes within the next few years.

FCC approval for the authorization of the new satellite launches was given unanimously, part of what commission chairman Charles Ferris described as an "open-entry policy" to provide satellite capacity to all who want it.

A separate report prepared by commission staff, for example, has recommended that direct-broadcast satellite television services should be subject to the minimum of regulation, in

the light of competitive pressures from conventional television stations and cable television networks.

The commission is also expected to publish for comment a proposal for squeezing more domestic satellites into geosynchronous orbits, perhaps by reducing the separation of higher-frequency satellites from four to three degrees of longitude.

Such a change, says FCC, would allow the launch of an additional six satellites which had been promised to the commission, but which could not be authorized because of the shortage of available slots.

The launch of the 20 new satellites is likely to result in an increase from 160 to 612 in the number of satellite communications channels available by 1985. Each channel can transmit 1,000 telephone channels or a single television channel.

The new satellites will also create the first major challenge to the giant American Telephone and Telegraph Company (AT&T) in the field of long-distance communication. Two competing companies — General Telephone and Electronics Corporation and the Continental Telephone Corporation — are building their own satellites. AT&T, General Telephone, Southern Pacific Communications Corporation and Hughes Communications will each be constructing their systems for the first time, each having been given permission to build three satellites and launch two.

Local newspaper production could be revolutionized by plans that the Gannett Company is said to be developing for a national newspaper produced in Washington, but beamed to the presses of some of Gannett's 82 local newspapers.

Satellite transmission is already used by the *Wall Street Journal* to print seven editions simultaneously throughout the country. Although no official confirmation has been given of Gannett's plans, it is reported that the initial investment would be about \$100 million; some commentators have pointed out that setting up a single newspaper distributed throughout the country would provide Gannett with the opportunity to become a major voice in national affairs.

Comsat's plans for a satellite-based television network are included in a set of comments which the company has submitted to FCC as part of its preparations for the Region 2 Administrative Radio Conference of the International Telecommunications Union, due to be held in the summer of 1983.

Four satellites would be used to cover the United States, each covering a different time zone, with the most westerly satellite also broadcasting to Hawaii and Alaska. The service would be operated by Satellite Television Corporation, a newly-created subsidiary of Comsat. The market for direct-broadcast satellites is expected to be tested with the modification of two existing spacecraft designs.

David Dickson

Ariane development

More trouble

Ground tests of the principal rocket motors of Ariane, putative launch vehicle of the European Space Agency (ESA), have shown brand-new oscillations in the burn — and set the test programme back by a second three months.

Ariane's first test launch in late 1979 reached the planned trajectory, but the second — in May 1980 — ended moments after lift-off with a high frequency oscillation and loss of pressure in one of the first stage engines. A look back at the data for the first launch showed the same oscillation, but at low amplitude; and a re-run of the data accumulated during ground tests showed the same fault.

In October this year, ESA announced that the oscillations (at 2,300 Hz) could be cured by improving tolerances in the manufacture of the injectors, devices like a watering-can rose that mix the fuel and oxidant. This hypothesis will be tested at the end of this month, when new high precision injectors are tried for the first time. Meanwhile a new oscillation has appeared — at 2,700 Hz — and this sets in even with injectors that previously showed no sign of the vibration. While the 2,300 Hz problem "can be regarded as rectified" says ESA hopefully, the 2,700 Hz oscillation "is the subject of thorough investigation and action which still needs some time to complete".

Burn oscillations (called "buzz" or "screaming") are the *bête noire* of liquid fuel engine design and can destroy the engine if they reach high amplitude. One solution is completely to redesign the injector; another is to set Helmholtz resonators, tuned to the critical frequency, into the side of the combustion chamber in such a way as to absorb the oscillation energy.

Solutions such as this, however, verge on complete redesign of the engine — but this may be necessary in the end. The engines were originally designed for the French sounding rocket Diamant (in which similar instabilities were met and solved), but they

were stretched to the limit to meet Ariane's specification of 61 tonnes-weight thrust. The next series of engines, for a bigger Ariane, is based on a different design.

So it is seriously being questioned whether Ariane can meet its first commercial commitments: to launch the French MARECS B and SERIO 2 marine communications and meteorological satellites (which would go up together) and Intelsat V F6. These satellites, and the ESA X-ray astronomy satellite EXOSAT, were originally due for launch in 1981.

For the moment, ESA hopes to have the third test launch in June 1981, and the fourth in the autumn, beginning the commercial series later in the year. If, however, as some fear, the Ariane engines are inherently "marginal", redesign and retesting could take at least a year, and lucrative launch contracts might be lost to American competition. The lead that Ariane appeared to be establishing over the space shuttle — beset by its own problems but now believed to be running smoothly — is narrowing fast.

Robert Walgate

Halley missions

NASA to go?

Washington

Can the new US Administration be embarrassed into mounting a mission to Halley's comet when it passes through the Solar System in 1986? The proposal has already been passed to president-elect Ronald Reagan from scientists in his home state of California. They point out that the European Space Agency, the Soviet Union and Japan all have plans for separate Halley missions — and that for the United States not also to go would be a major blow to national prestige.

Revised plans for a Halley fly-by were developed earlier this year at the Jet Propulsion Laboratory (JPL) in Pasadena, run for the National Aeronautics and Space Administration (NASA) by the California Institute of Technology (Caltech). Following last year's rejection by the White House of a request for funds to develop an ion-drive — for which money

Beyond Saturn

Saturn, as seen by Voyager 1 on 16 November, four days after the encounter. Some of the dark spoke-like ring features (*Nature* 4 December) are seen as bright patches due to scattering of sunlight from particles within them. Voyager 1 will be monitored to as great a distance as possible in an attempt to detect the boundary between interplanetary and interstellar space, where the "solar wind" becomes undetectable. Voyager 2 will reach Saturn on 25 August 1981, Uranus in 1986 and Neptune in 1989.



has since been allocated by Congress, but too little and too late — the revised mission would rely on a ballistic launch, limiting its ability to manoeuvre near the comet.

Many JPL scientists still think the mission would be worthwhile, a conclusion backed up by the mission's science committee, headed by Professor John Ververka of Cornell University. And the proposal has already been put to Mr Reagan by Caltech trustee Earle M. Jorgensen, a Los Angeles businessman and close friend of the president-elect.

Scientists at JPL are stressing the question of national pride, one aspect of the space programme that they feel has received relatively little support from the Carter White House. But other members of the space science community are worried that political support for a fly-by mission could prejudice funding for NASA's own favourite scientific project, the Venus Orbiting Imaging Radar (VOIR).

In an unusual move (considered by many a pre-election gambit to win Californian votes), the White House announced in October that it was planned to request funds from Congress for VOIR when the 1982 budget proposals are submitted in January — implying a rejection of the revised Halley plans.

From the scientific point of view, VOIR is reckoned to be more fruitful than the Halley intercept. Members of NASA's space science advisory committee have already stated that they find the possibility of losing VOIR and getting the Halley mission instead to be "extremely distressing".

In addition, however, to the financial attraction — the present Halley plans would make substantial use of instruments developed for the Voyager missions, and its total cost of \$250 million would therefore be only about half that of the proposed VOIR — JPL scientists point out that the chance to observe the comet is a "once in a lifetime opportunity".

Officials from the European Space Agency (ESA) are watching the new developments with some concern, since they feel that a NASA project might well overshadow the more modest goals of ESA's own Halley mission, Giotto, particularly since the latter may not be able to produce photographs of the comet comparable to NASA's.

A substantial NASA involvement in the European effort is now unlikely. Two weeks ago, ESA's space science committee decided to keep the launch a European affair, using the French Ariane rocket, and were therefore not interested in the US offer of using a Delta launcher in exchange for payload space.

US scientists will not be entirely excluded from the European mission. A number are listed as co-investigators in experiments submitted for inclusion in the spacecraft. And ESA is still negotiating terms for Giotto to make use of NASA's deep space tracking network.

NASA, however, is uncomfortable about accepting the unusual role of second fiddle, and supporters of a US mission are emphasizing this possibility in efforts to generate support. Adding the Halley mission to the 1982 budget for NASA would affirm to the world in a "spectacular and dramatic way" that US leadership in high technology stands unsurpassed at the frontiers of knowledge, according to a statement published in the *Congressional Record* by conservative Republican Senator Storm Thurmond. Another supporter is ex-astronaut Senator John Glenn.

If the mission is given the go-ahead, its planners will have a delicate balance to weigh up. Scientifically, more data will be obtained from intercepting the comet after it passes through its perihelion, since solar heating will stimulate the discharge of gas and dust particles. But politically an earlier intercept might be more attractive, as this would pre-empt both the European and the Soviet post-perihelion encounters.

Top NASA administrators are also said to be concerned about the risks of the mission — if the spacecraft is hit by a dust particle there might be no scientific return at all. Scientific and political priorities are therefore likely to meet head-on; nobody is predicting the outcome.

David Dickson

Polish universities Union snag

The drafting commission for the new Bill on Higher Education in Poland met for the first time two weeks ago, under the chairmanship of Professor Zbigniew Redich. The commission was set up by the Minister of Science, Higher Education and Technology, Dr Janusz Gorski. A few days previously, Warsaw University students had staged a sit-in in one of the main buildings of the university to express their lack of confidence in the minister.

The two events are not unconnected; the new bill will give Polish universities considerably more autonomy and a more democratic form of self-government which will include student participation. The two hundred students who occupied a hall in the university's Kazimierzowski Palace were protesting, first and foremost, against the difficulties they are encountering in registering the new Independent Students' Union (NZS). Since students are not "employees", they cannot register as an independent trade union under the Gdansk accords. The first draft of the ministry's "instruction" permitting registration of NZS proved unacceptable to the Warsaw students, although it had been worked out with the participation of NZS delegates. When the ministry negotiator suspended talks with NZS, on the grounds that he must first consult the (now greatly depleted) Socialist Students' Union, NZS considered that the ministry had broken

off negotiations unilaterally and began their sit-in.

After occupying the hall for two days, the students won what they described as a partial concession only — a new ministry instruction on registration to come into force on 20 December. This should allay at least one fear — that without legal status, NZS would be unable to participate in the new "self-government" system of the university. Demands for a national NZS delegate meeting on student problems and for the publication of their grievances in the press were not met.

Neither the concessions by the minister nor the conciliation efforts of the university rector, Dr Henryk Samsonowicz could persuade the students to end their sit-in. Ironically, in view of the fact that NZS cannot legally be a trade union, it ended its protest only in response to a call for restraint issued by the Independent Trade Union Confederation Solidarity.

Vera Rich

Fissile material Counting wrong

Are eleven kilogrammes of highly enriched uranium missing from Dounreay, the UK Atomic Energy Authority's (UKAEA) fast reactor research establishment in North Scotland? The authority cannot say. The amount appears in this year's tally of "materials unaccounted for" (MUF), and is something of an embarrassment. According to the UKAEA, the quantity is within measurement errors on a large throughput, but the throughput cannot be revealed for security reasons and the errors have not been calculated. So it is difficult to give significance to the figure.

All MUF figures announced last week lack error estimates because, an official said on Monday, "errors would have to be combined from many completely different processes and sources".

Nevertheless the 11 kg "loss" at Dounreay was near enough to a critical mass to give the authority pause. In the first three years for which MUF figures were announced, the MUFs at Dounreay for highly enriched uranium were +2.8 kg, +3.7 kg, and +0.3 kg, all indicating paper gains of the material. However, Dounreay was not reprocessing fuel for much of that period, and the present figures relate to the first year (since MUF accounting began) when fuel from the now closed Dounreay fast reactor was being reprocessed.

The figure amounts to a difference between the quantity of ^{235}U estimated to be on the site at the beginning of the year and the quantity estimated at the end, taking into account traffic on and off site. But much of the ^{235}U at the beginning was in the form of irradiated fuel rods, and so inaccessible to measurement. The amount in the rods was estimated using knowledge of their position and operating time in the

Dounreay fast reactor, and of the original composition of the rods. At the end of the year, some of these rods had been through reprocessing: here ^{235}U can be measured easily only at the end of the flow, and by neutron activation analysis in the cans (which are stripped off at the beginning of the process).

When the 11 kg accounting loss was discovered, great efforts were made to recalculate the quantity and check measurements. The new figure was different (the authority will not release it) but it was agreed to keep to the original accounting value. However, the difference was great enough to indicate that the uncertainties in the original estimate were of the same order as its value — about 10 kg. For example, the recalculated figure for ^{235}U in the original rods was 2 kg different from the first estimate.

However, still no calculation was made of the uncertainty, or "error", on the quantity. And this despite a statement a year ago by Dr A. G. Hamlin, director of the UKAEA Nuclear Materials Accounting Control Team, that "in order to determine whether MUF is really indicative of diversion of material, the safeguards authorities need to know, among other things, the expected level of errors. . . .".

Robert Walgate

US research planning

NIH reprieved

Washington

The National Institutes of Health (NIH) have avoided the threat of new legislation which would have placed much closer congressional surveillance and control on their research programmes, proposed earlier this year by the House of Representatives health and environment subcommittee (see *Nature*, 22 May).

After intense opposition from medical schools and research institutes — and despite the support of the Secretary of Health and Human Services, Mrs Patricia Harris — proposals put forward by Subcommittee chairman Henry Waxman were mostly dropped when the bill was agreed in final form by the House and the Senate.

Also dropped were equally controversial proposals put forward by Senator Edward Kennedy on the Senate side for a Presidential Commission on biomedical research priorities. Senator Kennedy had wanted to attach this to legislation which re-authorizes the operations of the National Cancer Institute and the National Institute for Heart, Lung and Blood, the only two of NIH's eleven institutes at present requiring such legislation.

NIH themselves opposed both innovations, but were told by Mrs Harris to sit on their hands. Opposition was subsequently led by groups such as the American Association of Medical Colleges, which argued that NIH should be given the greatest possible freedom to

decide on their research strategies.

NIH did not, however, escape completely unscathed. Included in the authorization bill are requirements for additional administrative arrangements covering diabetes research — a pet subject of retiring health subcommittee member Senator Richard Schweiker, expected to become Mr Reagan's Health Secretary — and for research on digestive diseases. NIH officials feel that one disadvantage of the changes proposed by Congressman Waxman is that regular re-authorization of NIH research funds makes it easier for such provisions to be attacked by individual congressmen.

David Dickson

European environment

Small gains

Brussels

Agreement has still not been reached by the Environmental Council of the EEC on two pieces of legislation: one concerning the prevention of industrial accidents and the other on the level of mercury discharge. However, new controls on whale imports, the recycling of waste paper, and wildlife habitation have been accepted.

At the meeting in Brussels on 12 December, the so-called "Séveso Directive" again foundered on the French intransigence over transfrontier notification procedures. It had been proposed that information on potentially hazardous plants near frontiers should be made available on a Community basis. The French would prefer notification of neighbouring states bilateral — an idea swiftly rejected by Belgium and Luxembourg. This disagreement is bound to step up the campaign against French nuclear reactors near the Belgium and Luxembourg borders.

The directive on mercury is also meeting resistance from France and Britain. The UK delegation insists that environmental quality objectives be used instead of standard emission limits to measure pollution. In the draft directive — the first to be based on the controversial directive on the control of discharges of dangerous substances into the aquatic environment — it was suggested that the United Kingdom stick to its use of quality controls while the other EEC states use the limit value approach. The French, however, object to this, feeling that pollution abatement costs should be the same throughout the Community. Both this and the industrial accident proposals will now go back to the Committee of Nine's Permanent Representatives for further discussion.

On the bonus side, from 1 January 1982 imports of commercial whale products will be banned. Would-be importers will have to apply for a licence not only for primary products but also for any goods (such as leather) which have been treated with a whale product. Denmark's insistence that

Greenland be excepted from this rule was accepted.

A recommendation obliging the nine member states to take steps to recycle waste paper and pulp was also adopted. Inks or glues which adversely affect the recycling process may fall victim should the legislation be seriously applied.

It was also agreed to finalize the Community's adherence both to the Strasbourg Convention (the Berne Convention) on the protection of European wildlife and their natural habitats and to the Geneva Convention on trans-frontier air pollution.

Jasper Becker

High-energy physics

New man, new style

The next director of DESY, the West German particle physics laboratory at Hamburg, will be 49-year-old Dr Volker Soergel, at present a member of the directorate of CERN, the European centre for particle physics in Geneva. Departing DESY director Professor Herwig Schopper announced the appointment during his leaving party at DESY last week.

Schopper moves to CERN in January as director-general; and Soergel will step into his shoes at DESY, even to the point of taking over his predecessor's professorship at the University of Hamburg. But from there on the similarities between the two individuals stop. Schopper had proved himself an extremely able politician in his period at DESY, but is less well known as a physicist; Soergel is a recognized expert on the weak decays of elementary particles, but has yet to learn the ways of Bonn.

However, he is by all accounts an extremely able administrator. According to one ex-CERN physicist, he could complete in ten minutes meetings that should have taken an hour — by resolving conflicts of interest beforehand. In this sense his style is ordered and Germanic. But at the same time, it is said, he can wax enthusiastically "and talk for hours about some harebrained scheme".

At DESY there are plenty of schemes, harebrained or not. A 30-beam synchrotron radiation laboratory opens in January; PETRA, the big electron-positron ring, is being pushed up to 22 GeV per beam in an attempt to find the elusive t -quark; superconducting magnets are being developed; and there are long-term plans for HERA, an 800-GeV proton on 35-GeV electron collider. But first, Soergel must find cash for DESY to pay its electricity bills (see *Nature* 4 December).

Being interested in the weak interaction, he will be keen to run PETRA at the energy where it can take data fastest on weak-electromagnetic interference, and check — in a hitherto impossible way — the Nobel-prize winning Weinberg-Salam unified theory of the weak and electromagnetic interactions. DESY physicists now

estimate that this requires experiments at 16 GeV per beam — not DESY's top energy, but enough to break the present budget if pursued for more than half 1981 running time. Given the cash, they say, they could have a result by May or June. It will be interesting to see if Professor Soergel can get it.

Robert Walgate

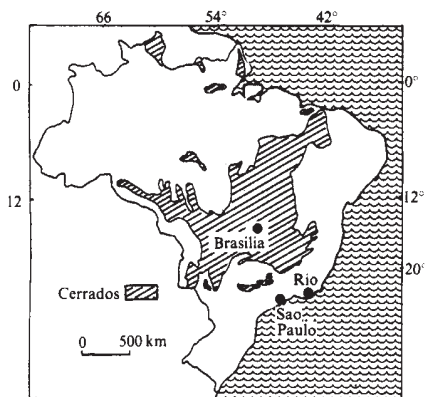
Brazilian agriculture

Soya beanfeast

Brasilia, November

In the scramble to develop its unexploited resources, officials say that Brazil is turning away from the Amazon Basin to an area of savannah, known locally as the Cerrados, covering about 180 million hectares mainly in the centre of the country. Attempts to open up the Cerrados, including the building of Brasilia, have so far produced only moderate results. But scientists have now convinced the government that the region offers great potential for agriculture.

The problem with savannah regions is that the very old soils have been leached of plant nutrients so only rough scrub supporting inefficient cattle grazing will grow. Aluminium concentrations in the Cerrados are also up to three times more than will be tolerated by most crops. Yet the soil structure and climate are good.



The problems of cultivation in the area are largely the responsibility of a government research institute, in the middle of the region, about 20 miles outside Brasilia. According to Dr Elmar Wagner, its scientific director, the aluminium and acidity problem can be solved by adding limestone, in quantities of roughly 2 tonnes per hectare depending on the crop to be grown. Fertility can be finally restored by adding approximately 150–180 kg of phosphate per hectare, and then maintained by the addition of nitrogen, phosphate, potassium, a trace of zinc and good soil and crop management.

Although Dr Wagner claims that agriculture can be maintained with less fertilizer than is commonly used in industrialized countries, fertilizing such a vast area would be a costly undertaking for Brazil, which has to import most of the raw materials for its fertilizer. So the search is

on for suitable nitrogen fixing crops.

Brazil has already experienced considerable success in breeding soya beans, now its main export crop, which fix all their nitrogen requirement. But attempts in 1979 to introduce nitrogen-fixing soya beans into the Cerrados from the southern coastal states were unsuccessful. Now a team at the Rural University of Rio de Janeiro claims to have discovered why.

According to Joanna Dobereiner, leader of the team, the sudden cultivation of newly-reclaimed tropical soils seems to increase *Streptomyces* fungi which produce antibiotics that kill off nitrogen fixing *Rhizobium* bacteria on the plants' roots. By isolating those strains of *Rhizobium* which survive and inoculating them into soya bean roots, she obtained good soya bean yields in a recent field trial in the Cerrados. She claims that by careful rotation of crops and by use of streptomycin-resistant *Rhizobium*, it may be possible to eliminate most of the need for artificial nitrogen fertilizer which accounts for 70 to 80 per cent of all fertilizer investment.

Most of the growth in farming in the Cerrados since 1975 has been in rough pasture for cattle raising. But Dr Wagner thinks that greater efficiency could be achieved by concentrating on grain crops and forestry. By cultivating most of the area available with improved methods he estimates that the total annual yield of grain crops could rise from 7.5 million tonnes now to 125 million tonnes and that of forestry from 15 million tonnes to 600 million tonnes. Meat production, he says, could be increased from 2.2 million tonnes annually now to 8.0 million tonnes even with pasture area reduced from 144 million to 80 million hectares.

With ambitious plans to cut oil imports further by using alcohol substitutes for petrol and vegetable oil substitutes for diesel, Brazil's need to open up more land to cultivation is becoming crucial. The problem was highlighted this year when Brazil was forced to import black beans from Mexico because some black bean land in the south had been turned over to alcohol production. The question now is whether the Cerrados could be used to grow food, leaving the traditional agricultural lands of the south free for sugar cane for alcohol, or whether the Cerrados itself could be turned over to sugar cane production.

Whatever the decision, the government will be faced with the problem of encouraging the growth of the Cerrados. It has been criticized in the past for encouraging the exploitation of land by wealthy southerners who seek quick profits and who have little interest in long-term development. A major challenge now will be transferring the results of research to the farmer and helping Brazil's underprivileged farming community in the north-east region to take part in the developments.

Judy Redfearn

Schmitt stars again

Washington

Eight years ago he was collecting rock samples on the Moon. This summer he was a key fund-raiser for presidential candidate Ronald Reagan. And next year he will be the central figure for science in the US Senate, with responsibilities ranging from the programme of the National Aeronautics and Space Administration (NASA) to the budget of the National Institutes of Health (NIH).

A graduate from the California Institute of Technology with a PhD in geology from Harvard, Harrison (Jack) Schmitt won a surprise election victory as Senator for New Mexico in 1976.



Astronaut (right) and Senator (left) Schmitt

Last week he was appointed to head the Senate Commerce Committee's subcommittee on science technology and space. The committee is responsible for overseeing the programmes of both NASA and — if Senator Schmitt has his way — the National Science Foundation.

Not surprisingly, Senator Schmitt has been a keen supporter of the space research programme. With Senator Adlai Stevenson, the subcommittee's present chairman, he has been strongly critical of the Carter Administration's "unimaginative" approach to space policy — and his appointment has been welcomed in the space science community.

More controversial is his selection as chairman of the Senate's labour, health and human services subcommittee — responsible, among other things, for the biomedical research budget of NIH. This post was expected to go to Senator Charles Mathias, a liberal Republican and an enthusiast for biomedical research. But Senator Mathias has now shifted his attention to the Judiciary Committee, where he will counterbalance that committee's new chairman, conservative Senator Strom Thurmond.

Senator Schmitt's attitude towards this area of research is little known. Although a member of the present subcommittee, he has played little part in its deliberations on the NIH budget. As a geologist, he is expected to favour support for basic science; as an astronaut, NIH officials fear he may be drawn towards the high technology aspects of medical care, rather than its social and environmental dimensions.

David Dickson

CORRESPONDENCE

Nuclear cheap?

SIR—May I comment belatedly on Professor Jeffery's letter on nuclear costs (*Nature* 23 October, p.674). It is pointless trying to recalculate electricity generation costs on the basis of arbitrary rules. If, as Professor Jeffery believes, capital costs are inadequately reflected, does he propose to recalculate the current costs of coal and its transport in terms of capital expended in the nineteenth century? How does he propose to take account of the fact that if there had been no nuclear contribution the additional fossil fuel supplies would have had to come from expanded output at the marginal and most expensive sources? The costs published by the Central Electricity Generating Board (CEGB) for generation from Magnox nuclear stations and coal and oil fired power stations of similar age and load history, show that on a perfectly proper accountancy basis, including allowance for costs yet to be incurred, the price we would be paying for electricity would be higher now had fossil stations been preferred in the past.

The industry has stressed, however, that this alone is not a guide to future investment, any more than it would be if Magnox fuel costs increased more rapidly than coal and reversed the current cost position. Incidentally Professor Jeffery's projection of fossil fuel prices in terms of the sum of coal and oil prices compounds changes in fuel mix and offsets rising coal prices with the decline in real oil prices following the major price excursion in 1973. It conceals the 60 per cent real increase in coal prices which occurred between 1970 and 1979.

Fuel cycle costs for advanced gas-cooled reactors or pressurized water reactors are different from Magnox in that they refer to oxide fuels with higher burn up for reprocessing in THORP. Estimated future fuel costs are based in part on the expected cost of this plant. Electricity costs for stations yet to be constructed are calculated in constant money terms on a leveled cost basis by summing expected future discounted expenditures and dividing by expected discounted output. Our calculations, which may differ in detail from those of CEGB, suggest that even if coal costs were held steady and uranium prices allowed to rise (both contrary to present trends) nuclear should retain a small cost advantage.

P.M.S. JONES

UK Atomic Energy Authority,
London SW1, UK

Biospheric works

SIR—I have just read your article concerning Unesco in the 6 November issue of *Nature*. While I shall refrain from commenting upon the overall content of the article—although for instance we do not know here of any "new technological order" which would have been approved by the recent Belgrade session of our General Conference—there is one point in your text which particularly puzzles me. You state that our programme on Man and the Biosphere (MAB)—in which many scientists and institutions actively participate all over the world—is a source of "largely empty generalizations" and that "people and member governments increasingly ask whether it can be worth its cost".

A few years ago I outlined this programme to your readers in a lead article which appeared in *Nature* of 17th July 1975. The programme was still only emerging from its preparatory phase. It is now fully operational, with close to a thousand very concrete research projects under way in some 75 countries. Around 200 biosphere reserves have been established under the programme in 50 countries. Networks of integrated pilot research, training and demonstration projects have been set up in the humid tropics and in the arid grazing lands. Cities like Rome, Frankfurt or Mexico are being studied as urban systems under MAB. The annual cost of the programme to Unesco is approximately \$2 million only. This catalytic amount has attracted—mainly from industrialized countries—some \$50 millions of extra-budgetary support. In the aggregate it is estimated that national commitments to MAB run at about \$200 millions per year. It would indeed be strange that research organizations would spend such amounts for applied ecological studies leading only to empty generalizations. And everyone present at Belgrade can testify that the programme received enthusiastic support from all governments represented, whether developed or developing.

As a final point, let me indicate that MAB will be 10 years old this year and that we are organizing next October jointly with the International Council of Scientific Unions a conference-exhibit to review the progress of MAB, to evaluate its achievements and shortcomings and to make recommendations for its future. Any constructive criticism which *Nature* and its readers would wish to make would be most appreciated. But sweeping unsubstantiated statements are of no use in this context.

M. BATISSE

Unesco, Paris, France

More on museums

SIR—May someone who is not by profession a scientist join in your controversy about "Museum pieces"? For like so many other non-scientists I do have a strong interest here, and that an interest of the kind which surely ought, in the present case, to be overriding. That is to say, I am a devoted museum goer myself, with children who also need the help of museums in their education. What we consumers of museum services are entitled to expect from our national institutions is a clear, vivid, fair and balanced presentation of all available materials. When there is fundamental disagreement on any issue of substance between the best qualified professional experts, then we are entitled to be told that such disagreement exists and what the main points at issue are.

Dr L.B. Halstead has made a very serious charge (*Nature* 20 November, p.208) that in the Natural History Museum the exhibits of dinosaurs and of fossil man are now presented in terms of a new theory which none of his critics has yet even suggested represents the consensus of the experts. Nor, it would appear, have those responsible for making up these exhibits and compiling the accompanying booklets done anything to explain and to stress that theirs is only one

view, and that probably a minority view, among those who are best qualified to judge. I would emphasize again that I make no pretensions whatsoever to any such special expertise. But it requires no expertise at all to recognize that Dr Halstead's charges are very serious indeed, and that none of the critics whose letters you have so far published has made any attempt to show that the situation is not in fact as scandalous as Dr Halstead's original letter suggested that it is.

ANTONY FLEW

Department of Philosophy,
University of Reading, UK

SIR—in response to correspondence from L. B. Halstead (*Nature* 20 November, p. 208), I feel obligated to point out that while science is generally regarded as the outgrowth of processes of the cerebral cortex, Halstead's criticisms of cladistics seem to emanate more from the anterior pituitary.

First, although phylogenetic systematics (cladism) and the punctuated equilibria mode of evolution are largely complementary in their theoretical structure, at least one articulate spokesman for punctuated equilibrium is not a cladist, and therefore it is incorrect to lump the two schools together.

Next, although I am as firmly convinced of gradualism and evolutionary systematics as anyone, it is academic vigilantism to deny the cladists and punctuationalists their day in court. Those views are propounded by extremely competent scholars, whom I perceive to be fundamentally committed to the advancement of our understanding of the history and dynamic processes of life on Earth. This is a different goal (in fact, a directly opposite goal) from that of the creationists, and it is highly inappropriate for Halstead to analogize between their motives.

As for Halstead's innuendo that punctuational evolution is a Communist plot, I confess that I am never quite sure when British people are kidding, but I hope that Halstead was. Gould and Eldredge did spend one page of a thirty-five-page paper (*Paleobiology* 3, 115; 1977) relating their punctuational model of evolutionary change to the more general views of change espoused by Marx. Halstead's criticisms, however, do not address either the issues or the data; they instead take the form: (1) Marxism is wrong; (2) therefore, everything a Marxist says is wrong. This is obscurantism of a most pernicious sort, and side-steps the critical question of whether or not cladism or a punctuational mode of evolution has anything important to contribute to our understanding of the world.

When I see the maxillary and mandibular remains of the Miocene and Pliocene hominoids, I see (along with Halstead, I suspect) anagenesis, migration, allopatric speciation and gradual genetic divergence—without necessary recourse to a punctuational scheme. I do not see, however, how the vituperative essay of Halstead's sheds any new light on any relevant issues. Science should proceed through the critical analysis of ideas and explanations by recourse to the facts at hand: processes of logic and reason. Let us keep our hormones out of it.

JON MARKS

Department of Anthropology,
University of Arizona, Tucson

NEWS AND VIEWS

Climate and variability in the solar constant

from Douglas Gough

AT one time or another almost every fluctuation in the Earth's climate has been attributed by someone to a variation in the solar constant. In many cases alternative, more convincing explanations have been found. For others there is mounting evidence that Milankovich's idea is correct: that variation in terrestrial insolation arising from changes in the elements of the Earth's motion is an important controlling factor. But there remain components of the climatic variation that are otherwise unexplained, and so it has recently become fashionable to study the implications of changes in the solar constant that arise from intrinsic variations in the solar energy flux. This has occurred partly because observational techniques now make it feasible to measure the small fluctuations that many meteorologists and climatologists believe to be sufficient to produce deviations in our weather.

At a recent workshop* issues concerning the intrinsic variability of the Sun were discussed. The principal aim was to consider observations of the solar luminosity and radius, and to determine to what extent solar phenomena are correlated with changes in terrestrial climate.

The meeting began with a review by G. North (Goddard Space Flight Center) of climate modelling. This very difficult subject is still in a primitive state and any conclusions drawn from it must be viewed with caution. In 1973 the response of a rather sophisticated climate model (that devised by Tzvi Gal-Chan and S. Schneider at the National Center for Atmospheric Research) was investigated to see if a reduction in the solar constant of a few per cent would be sufficient to cause an ice age. The decrease was found not only to bring about an ice age but also produced a result that had not been anticipated: when the solar constant was restored to its original value, or even to a value somewhat higher, the ice would not melt. Astrophysicists believe firmly that the solar luminosity has evolved gradually during the past 4.7×10^9 years from about 70 per cent of its present value, and thus the implication of that climatological experiment appeared to be that the Earth must always have been covered with ice. This is evidently not the

case. Therefore it seems inevitable that the climate model must be inadequate. Presumably the Earth has some stabilizing mechanism which operates only on a long time scale and which has not been taken into account. The Earth must be stable to the gradual evolutionary variation of the solar constant, but perhaps not to fluctuations on shorter time scales.

According to North, the situation has not changed since that experiment was performed. There have been various attempts to accommodate the evolutionary rise in the solar luminosity, by supposing the composition of the Earth's atmosphere to have changed such that the resultant variation in the atmospheric greenhouse effect cancelled the effect of the changing solar constant. But this variation has not been coupled either to the flux of incoming radiation, or to the state of the climate, so it appears to be quite accidental that the Sun and the Earth have kept in step. If one accepts that this is unlikely, one is left with what has recently been called the faint Sun paradox: either the astronomers are wrong or the climatologists are wrong; and if the latter is the case, either a long-term climatic stabilizing mechanism must exist or the climate is not really sensitive to changes in insolation on any time scale.

The evolutionary rise in the solar luminosity seems difficult to question, as A. Endal (Louisiana State University) emphasized. It is an apparently straightforward consequence of basic atomic and nuclear physics, gas kinetics and Newtonian gravitation. Therefore the astrophysical conclusion is robust. The sole caveat one might envisage is that perhaps the constant of gravitation (G) is decreasing with time, as some cosmologists have suggested. The rate by which G must decrease to prevent a substantial variation in the solar constant is $-\dot{G}/G = 1 \times 10^{-11} \text{ y}^{-1}$, which is below the present level of detectability. However, in the foreseeable future planetary ranging will no doubt enable measurements to be made.

On the other hand, the climatic balance appears to be rather delicate. A satisfactory self-contained theory does not exist, and the toy models, as North called them, rely on parameters which are determined from present-day empirical evidence and which are assumed to remain constant for all time. Evidently if the dependence of those parameters on the state of the climate were known and could be incorporated into the toy models, a drastic change in the stability

characteristics of those models could result.

To determine whether climate responds to small changes in the solar output, attempts have been made to correlate its variation with the solar cycle. This is a difficult task, since the response is only a small component of total climatic variation. Extensive work has been carried out by J.M. Mitchell, C.W. Stockton and D.M. Meko (in *Solar-terrestrial influences on weather and climate* p.125, eds McCormac, B.M. & Seliga, T.A., Reidel, Dordrecht, 1979), who have obtained an index of the average drought in the US since 1700 from tree-ring analysis. They have found a strong though imperfect correlation with the 22-year cycle, which agrees better with the actual sunspot record than with a regular oscillator with the same mean period. P.R. Bell (IEA) pointed out that an even better correlation exists between the drought record and a combination of the solar cycle and the lunar tides. The agreement is impressive, though its significance was challenged by Mitchell (NOAA, Silver Spring). According to M. Stuiver (University of Washington, Seattle) there is evidence for solar cycle variations in the ^{14}C record, but there do not seem to be global climatic changes associated with the Maunder and Spörer minima (see *Nature* **286**, 868; 1980).

Analyses of this kind have a bearing not only on climatology but also on the physics of solar variability. If an unambiguous relation between climate indices and certain aspects of solar variation could be established, it would be possible to attempt heliochronology from proxy records. Then one might be in a position to establish whether the solar cycle is regulated by an oscillation in which the entire Sun participates, or whether it is a phenomenon that is confined to the convection zone and its immediate environs.

A key to this problem may be the relative changes in the luminosity L and radius R , measured by the parameter $W = \Delta \ln R / \Delta \ln L$, where Δ represents the small change associated with the cycle. It was agreed at the workshop that if the cycle were a purely superficial phenomenon, then $W \approx 0$. If, on the other hand, it is deep-seated, the magnitude of W is of

*The workshop on solar constant variations was held at the Goddard Institute for Space Studies in Greenbelt, Maryland from 5 to 7 November, 1980. An extensive review of the meeting will be published by S. Sofia in due course.

Douglas Gough is in the Institute of Astronomy and the Department of Applied Mathematics and Theoretical Physics, University of Cambridge.

order (though rather less than) unity. Moreover, it appears that $W > 0$ if surface variations result from perturbations throughout or at the base of the convection zone, as E.A. Spiegel and N.O. Weiss (*Nature* **287**, 616; 1980), for example, have argued, and $W < 0$ if the perturbing process is confined to the core, as R.H. Dicke (*New Scientist* **83**, 12; 1979) upholds. Measurement of W is thus of great interest.

A whole day of the workshop was devoted to measurements of the solar constant. Rocket and balloon flights have not been frequent enough to gather much reliable information about trends, and ground-based observations have been plagued by interference from the Earth's atmosphere. In a critique of the day's proceedings, C. Fröhlich (Physikalisch Meteorologisches Observatorium, Davos) concluded that the data were consistent with no variation, but that a secular decrease in L from 1966 at the rate of 0.1 per cent per decade is not ruled out. It must be appreciated that this apparent trend is merely the result of linear regression through data with large gaps. Recent measurements of visible and ultraviolet sunlight from the Nimbus and SMM satellites suggest a decrease in L at the rate of about 0.1 per cent per year since 1976, the year of the last sunspot minimum. If all the data are taken into consideration and the possibility of some observational inaccuracy is accepted, one might also see with an eye of faith the hint of a variation associated with the sunspot cycle, with amplitude a few tenths per cent and such that luminosity maxima occur at sunspot minima.

There is another interesting luminosity variation which has a lesser bearing on the main issues confronted at the workshop, and which was agreed by all concerned with the measurements to have occurred. This is an occasional decrease of order 0.1 per cent which persists for about 10 days. On the two occasions on which it has been observed there was a very large sunspot group on the solar disk. Thus it appears that sunspots can reduce the total heat flux through the photosphere for intervals comparable with the turnover time of the larger eddies in the convection zone. This indicates that the sunspot phenomenon is deep-seated, and not confined merely to the supergranular layers as some workers have maintained.

To complete the measurement of W one must obtain R . This might be inferred from the effective temperature via the black body radiation law, once the luminosity is known. Spiegel and Weiss (loc. cit) have used estimates of temperature variations made by W.C. Livingston (*Nature* **272**, 340; 1978) from the CI 5380 spectrum line. However, Livingston (Kitt Peak Observatory) reported at the workshop that subsequent analyses of other lines, formed higher in the solar atmosphere, do not agree with the earlier conclusion. It appears that the atmospheric temperature

gradient has decreased between 1976 and 1980, but how this relates to the photospheric radius is unclear.

Direct measurements of the solar radius were discussed at some length. Most of this concerned trends on a time scale of centuries, which one might imagine to have been a relaxation from the Maunder minimum. There is a good deal of scatter in the data, and this no doubt has arisen in part from observational error. Eddy (High Altitude Observatory) reported a re-analysis of the Greenwich meridian circle data, and stands by his earlier conclusion (Eddy, J.A. & Boornazian, A.A. *Bull. Am. astr. Soc.* **11**, 437; 1979) that the Sun appears to be shrinking, though the rate has been revised downwards. The estimate is now about 0.1 per cent per century, in rough agreement with the values reported by D. Dunham (Computer Science Corporation, Silver Spring) from eclipse analysis, and by B.J. LaBonte and R. Howard (Mount Wilson Observatory) from Mount Wilson photographs. It was noted that this disagrees with I. Shapiro's null result (*Science* **208**, 51; 1980) obtained from Mercury transit measurements. Ancient transit measurements, like the Greenwich data, were dependent on observers' judgement, Eddy said, and it is hard to assess the accuracy of the results.

With this remark in mind it is perhaps not surprising that Parkinson *et al.* (*Nature* **288**, 548; 1980) have concluded that Eddy and Boornazian's result is not a property of the Sun, but is simply a measure of the variation in not only the observers' opinions but also the seeing conditions and instrumentation. Parkinson *et al.* rediscuss the meridian circle measurements in this light, and also consider Mercury transit and solar eclipse data. They find no convincing evidence for a long term trend, though they do see the hint of a variation on a time scale of about 80 years. It has long been known that this is a time scale

associated with the sunspot cycle (Wolf, R. *Astron. Mitt. Zürich* **14**, 1; 1862). It characterizes an irregular amplitude modulation of a kind common to a large class of nonlinear oscillators. It is not unlikely, therefore, that the seat of the modulation is intimately connected with the mechanism of the 22-year oscillation itself. Therefore the phase of the modulation relative to the radius variations is not without interest. Comparing figure 3 of Parkinson *et al.* with the sunspot data (for example, Waldmeier, M. *The sunspot-activity in the years 1610–1960* Schulthess, Zürich, 1961) suggests that maximum radius occurs at about the time when the amplitude of the sunspot cycle is at a minimum.

On a shorter time scale, LaBonte and Howard have deduced from Mount Wilson photoelectric measurements that a decrease in radius of about 0.01 per cent has occurred since 1974. During most of that time sunspot activity was increasing. T. Duvall (Goddard Space Flight Center) has made measurements of R using the vacuum tower at Kitt Peak National Observatory and H.A. Hill's (University of Arizona) finite Fourier transform definition of the solar limb. At the workshop he reported a decrease of 0.02 ± 0.02 per cent since last December. Hill plans to improve the measurement, and hopes to achieve an accuracy of one part in 10^7 .

In summary, we cannot really be sure how either L or R is varying on time scales of decades or centuries. Measurements of radius variations over centuries are subject to a good deal of uncertainty, and direct luminosity measurements do not exist. But if one accepts the hints from the most recent observations, one may be tempted to infer that perhaps there are variations in L and R over the sunspot cycle, with maxima in both L and R occurring at sunspot minima. If that is indeed the case, then $W > 0$. □

Development of spinal sensory systems

from T.M. Jessell and M. Yamamoto

CLASSICAL studies on the development of the spinal cord have focused on the motor neurone and its peripheral target, and the part played by synapse elimination and cell death in the modelling of connections¹. Yet motor neurone death represents only the first in a series of changes which involve the elimination or reorganization of sensory and spinal neurones and which are

T. M. Jessell and M. Yamamoto are in the Department of Pharmacology, St. George's Hospital Medical School, University of London.

probably not completed until well after birth. Staining techniques that make it possible to trace individual afferent nerves have shown that sensory fibres that can be distinguished on the basis of their peripheral receptive properties also have unique and characteristic terminal arbors in the central nervous system. But how such specificity is generated is only just beginning to be understood.

In the developing rhesus monkey, for example, sensory fibres enter the dorsal horn of the spinal cord during the first

quarter of the gestational period and form well defined synaptic connections with cells occupying the most superficial border of the dorsal horn². Three weeks later, however, these superficial neurones have degenerated, to be replaced by migrating cells that represent the future laminae I and II neurones. This loss of central targets may contribute to the degeneration of sensory neurones that is known to occur during normal development. But not all the sensory axons deprived of their post synaptic targets necessarily degenerate; others may survive to establish new and more stable synapses with superficial dorsal horn neurones².

Moreover, remodelling is not only a question of the elimination and re-establishment of afferent and efferent connections; the spinal inter-neurones themselves undergo extensive morphological changes as they mature. In new-born mammals the neurones in the marginal zone and substantia gelatinosa have stubby, radial dendrites that lack any defined orientation. During the first two weeks after birth they shed some dendrites and elongate others to arrive at their final characteristic morphology³. The cues for neuronal elimination and dendritic reorganization in the dorsal horn are unknown and the contribution, if any, of afferent input may be difficult to determine. In the cerebellar cortex where similar problems have been examined in greater detail, the development of Purkinje cell dendrites appears to be partially dependent on intact climbing fibre or granule cell inputs^{4,5} although even in the absence of presynaptic elements, a relatively elaborate dendritic arbor can still be generated.

These data raise interesting questions about the mechanisms by which the morphology of individual neurones is changed. At present, the only evidence bearing on this issue is ultrastructural. The additional membrane necessary for elongation of dendrites in the spinal cord, for example, seems to be generated by the fusion of vesicles with existing dendritic plasma membrane⁶; and degenerating spinal dendrites first become apparent as beaded processes containing numerous vesicles that eventually fuse to form cavities. Extraneous synapses on motor-neurones, on the other hand, seem to be removed by a rather more involved procedure. During the early post-natal period the number and distribution of synaptic terminals on motor-neurones decreases by about 50 per cent⁷. In a series of beautiful ultrastructural studies, Ronnevi⁸ has demonstrated that the phagocytosis of some synaptic terminals on motor-neurones may be undertaken by the motor-neurone itself. Motor-neurone cell bodies can be seen to contain spherical presynaptic inclusions that have double-walled membranes and are packed with synaptic vesicles. The motor-neurone is highly selective in the class of synaptic

terminal it engulfs, and other terminals are removed only by glial cells. Very similar double-walled vesicular structures have been noticed in embryonic mouse spinal cord⁹ and in the immature cerebellar cortex¹⁰ suggesting that this form of membrane remodelling may in fact be quite common in early development.

While investigations on the specific morphology and connectivity of sensory neurones have only relatively recently been attempted, it has not been until the past year that serious investigations on their specific chemical differentiation have become possible. The difficulty was that no sensory neurotransmitter had been identified, and the breakthrough has come with the discovery of substance P, somatostatin, cholecystokinin (or one of its shorter fragments) and vasoactive-intestinal polypeptide within small dorsal root ganglion neurones¹¹. While it is still unclear what part these peptides may play in synaptic transmission or modulation, they have provided several new markers with which to examine the differentiation of sensory neurones. Kessler and Black¹² have shown that the substance P content of rat sensory neurones is markedly increased after post-natal administration of nerve growth factor (NGF), which is already known to be essential for the early development of dorsal root ganglion cells. Conversely, neonatal administration of anti-NGF antibodies depletes substance P from the cell bodies and terminals of immature sensory neurones¹³. Using rhodamine-conjugated NGF¹⁴ it has been possible to show that NGF binds to receptors clustered at the tips of growing axons and is internalized and transported back towards the cell body of the neurone. Whether the internalized receptor complexes play any part in the actions of NGF on sensory neurones is, however, not yet known. Nor is it clear in these circumstances whether NGF is enhancing the survival of neurones already committed to the synthesis of substance P, or inducing substance P synthesis in neurones which would otherwise not have synthesized it. Two recent developments may make it

possible to answer this kind of question. First, there is evidence from experiments *in vitro* for two subpopulations of dorsal root ganglion neurones, only one of which depends for survival on NGF: the other will die in the presence of even a high concentration of NGF, but can be rescued by a factor derived from cultured glioma cells^{15,16}. The generation of purified subpopulations of sensory neurones in culture on the basis of their growth-factor sensitivity may thus soon be feasible, and should help to resolve the question of whether the effects of NGF on peptide synthesis are due to selection or induction of cells.

The second important development is the discovery that 'factors' derived from non-neuronal cells can influence peptide synthesis by cultured sensory ganglion cells. Cultures of chick sensory neurones grown in the absence of other kinds of cells contain eight times as much substance P as somatostatin¹⁷. In the presence of ganglionic non-neuronal cells, the same number of neurones make six times as much somatostatin as substance P. This dramatic switch (analogous to the adrenergic-cholinergic switch reported for autonomic neurones *in vitro*¹⁸) can be reproduced by adding medium conditioned by non-neuronal cells to 'neurone-alone' cultures, but is independent of NGF-like material. This suggests that early in development a given sensory neurone may have the capacity to synthesize and store more than one peptide. Final determination of peptide expression is probably induced, by factors other than NGF, at a later stage, since somatostatin and substance P are present in separate populations of small dorsal root ganglion neurones in adult animals¹⁹. Cholecystokinin and vasoactive-intestinal polypeptide are also found in some sensory neurones so the choice open to a single neurone may be greater still.

In adult animals, peptides in sensory neurones have been assumed to function in synaptic transmission. They may however also have a trophic role during development, as has been suggested for the conventional neurotransmitters noradrenaline and 5-hydroxytryptamine. Destruction of noradrenergic neurones in the locus coeruleus of neonatal rats modifies the postnatal arborization of pyramidal cell dendrites in the neocortex²⁰ while inhibition of 5-hydroxytryptamine synthesis in embryonic rats delays the development of neurones in regions normally receiving a 5-hydroxytryptamine input²¹. Peptides released from immature sensory neurones as they grow into the spinal cord could then regulate the differentiation of surrounding neurones. The use of anti-NGF antibodies or capsaicin to deplete sensory neurones of substance P would provide one method of analysing the role of substance P and possible other peptides in the development of spinal neurones. □

1. Purves, D. *Nature* **287**, 585 (1980).
2. Knyihar, E. *et al. Science* **202**, 1206 (1978).
3. Gobel, S. J. *Comp. Neurol.* **180**, 375 (1975).
4. Bradley, P. & Berry, M. *Brain Res.* **142**, 135 (1978).
5. Sotelo, C. *et al. Brain Res.* **98**, 574 (1975).
6. Falls, W. & Gobel, S. F. *Comp. Neurol.* **187**, (1979).
7. Conradi, S. & Ronnevi, L. O. *J. Neurocytol.* **6**, 195 (1977).
8. Ronnevi, L. O. *Brain Res.* **162**, 189 (1979).
9. Vaughn, J. E. & Sims, T. J. *J. Neurocytol.* **7**, 337 (1978).
10. Eckenhoof, M. F. & Pysh, J. J. *J. Neurocytol.* **8**, 623 (1979).
11. Hokfelt, T. *et al. Nature* **284**, 515 (1980).
12. Kessler, J. A. & Black, I. B. *Proc. natn. Acad. Sci. U.S.A.* **77**, 649 (1980).
13. Orten, U. *et al. Nature* **287**, 158 (1980).
14. Levi, A. *et al. Proc. natn. Acad. Sci. U.S.A.* **77**, 3469 (1980).
15. Barde, Y. A. *et al. Proc. natn. Acad. Sci. U.S.A.* **77**, 1199 (1980).
16. Barde, Y. A. *et al. Nature* **274**, 818 (1978).
17. Mudge, A. *Soc. Neurosci. Abstr.* **5**, 584 (1979).
18. Patterson, P. H. A. *Rev. Neurosci.* **1**, 1 (1978).
19. Hokfelt, T. *et al. Neuroscience* **1**, 131 (1976).
20. Maeda, T. M. *et al. Brain Res.* **70**, 515 (1974).
21. Lander, J. M. & Krebs, H. *Dev. Neurosci.* **1**, 15 (1978).

Successes and failures in human *in vitro* fertilization

from Alexander Lopata

THREE infertile women have given birth to normal infants following *in vitro* fertilization (IVF) and embryo transfer (ET). In each of these women a single embryo had been implanted into the uterus during the early luteal phase of a natural menstrual cycle on one or several occasions^{1,2}. Three other women conceived following ET but spontaneously aborted the fetus. One miscarried a 69XXX triploid fetus at 7 weeks gestation, a second miscarried an anatomically normal male fetus whose karyotype showed pleomorphism at 15D and a large Y chromosome at 21 weeks³, while the third miscarried a completely normal male fetus at 20 weeks following chorioamnionitis caused by an anaerobic Gram-negative bacillus.

The four pregnancies reported by Edwards and Steptoe¹ resulted from 32 embryo transfers and the two pregnancies reported by Lopata *et al.*² were obtained after 14 embryo transfers in 1979. After a number of changes were introduced in their laboratory the latter group carried out 10 additional embryo transfers by mid 1980 but have not obtained further pregnancies. Table 1 was compiled from the combined data and shows that at present the average efficiency of human embryo transfer is approximately 11 per cent whereas the probability of having a child following embryo transfer is three per cent. In contrast to this, live births have not yet occurred following transfer of one or more cultured embryos into the uterus of women whose ovaries had been stimulated with gonadotrophins or with Clomiphene citrate and hCG. Moreover, the efficiency of embryo transfer appears to be very low in these subjects since only three abortive pregnancies resulted from more than 77 embryos transferred by Edwards *et al.*¹ and biochemical evidence of early pregnancy was obtained in only three subjects

following 48 embryos implanted by the Melbourne group.

It should be noted that the computed success rate shown in Table 1 would be much lower if it were calculated on the basis of the total number of subjects monitored (treatment cycles) rather than on the basis of the number of embryo transfers carried out. Edwards and Steptoe monitored 78 menstrual cycles to obtain the 32 embryo transfers, whereas the Melbourne group monitored 69 subjects in 1979 and 63 in 1980. Thus, out of 210 women participating in the IVF and ET programmes, six became pregnant and of these, three delivered normal infants. The large wastage ensues from one of the following: failure to detect the distinct surge of luteinising hormone (LH) which is used for timing oocyte collection; ovulation occurring shortly before the ripe follicle is aspirated; the ovary containing the dominant follicle being inaccessible at laparoscopy; failure to recover an egg when the preovulatory follicle is aspirated; failure of fertilization or normal cleavage of the egg; failure of the embryo to establish pregnancy after it is transferred into the uterus.

Timing the collection of a preovulatory egg on the basis of 3-hourly urinary LH assays to detect the beginning of the gonadotrophin surge, which Edwards and Steptoe and the Melbourne group have used, is complicated by the following considerations. An oscillation in LH excretion is frequently detected before a distinct surge begins. The slope of the gonadotrophin surge and the peak value obtained may vary considerably between patients and often between menstrual cycles in the same patient. As these variations are inherent to the 3-hourly urinary LH assays, the criteria used to judge the start of the surge must be clearly defined to allow comparison of the time intervals (from start of surge to oocyte collection) selected by different groups of investigators.

Timing may begin from the baseline after the LH surge is detected, or from the mid-point of the first 3-hour interval when a sequential increase in LH begins, or from when the level of urinary LH excretion has increased by more than one, or two, standard deviations above the average baseline value. When timing began from the mid-point of the first 3-hour interval at which a sustained increase in urinary LH was detected, it was found that eggs recovered at 24 hours either failed to fertilize *in vitro* or underwent abnormal fertilization, whereas eggs recovered after 26, 28 and 30 hours yielded morphologically normal embryos. The two pregnancies reported by our group resulted from eggs collected 28 hours after the onset of the urinary LH surge. Steptoe *et al.*³ have reported that their first two pregnancies resulted from eggs collected at 26 hours after the LH surge and the second two pregnancies from eggs aspirated after 24 hours. However, the criteria used for starting the timing of the interval for oocyte collection have not been defined by this latter group.

The time sequence of intrafollicular maturation of preovulatory eggs in women ovulating spontaneously needs to be further investigated. Using ultrasonic scanning of the preovulatory follicle, commencing at 27 hours after the onset of the LH surge and rescanning several times per hour until ovulation occurred, de Crespigny *et al.*⁴ found that the interval from the start of the urinary LH surge to ovulation varied from less than 27 hours to 35 hours in nine women. Although this established that the time for follicular rupture varied by up to 8 hours between subjects it is now important to determine whether intrafollicular oocyte maturation requires a more constant time interval and is therefore independent of the actual time of ovulation. If, for example, preovulatory oocytes mature 24 hours after the LH rise, the follicle should be aspirated 25 to 26 hours after the onset of the surge to avoid spontaneous ovulation which occurs in about 20 per cent of subjects by 28 hours. On the other hand if meiotic maturity is incomplete, maturation of the egg in culture for several hours before insemination may improve results. Conversely, even brief periods of culture may be detrimental to mature human eggs, particularly if they resemble mature rabbit eggs which after 2 to 3 hours of ageing in culture remain fertilizable, but yield significantly fewer viable embryos⁵.

In vitro fertilization of human eggs has been reported to occur in modified Tyrode's solution formulated by Bavister⁶

Table 1 The current human embryo transfer efficiency and the probability of having a child by *in vitro* fertilization and embryo transfer

<i>Oldham/Cambridge group</i>			
Embryo transfers	32		
Pregnancies	4		
<i>Melbourne group</i>			
	1979	1980	
Embryo transfers	14	10	
Pregnancies	2	0	
Combined Transfer Efficiency	$= \frac{\text{Total pregnancies}}{\text{Total embryos transferred}} = \frac{6}{56} = 10.7\%$		
Probability of having a child	$= \frac{\text{Live births}}{\text{Embryos transferred}} = \frac{3}{56} = 5.4\%$		

Alexander Lopata is in the Department of Obstetrics and Gynaecology, University of Melbourne, and the Reproductive Biology Unit, Royal Women's Hospital, Carlton, Victoria, Australia.

and adapted by various investigators^{7,8,9}, in media described by Biggers, Whitten and Whittingham¹⁰, and by Hoppe and Pitts¹¹ as well as in modified Ham's F10 medium². Features common to these media include: a bicarbonate content which has been adjusted to maintain a pH between 7.4 and 7.6 in an atmosphere of 5 per cent CO₂; an osmolality adjusted to between 280 and 285 mOsmol/kg; supplementation with either crystalline albumin (bovine or human), or with inactivated human serum, both of which are believed to favour *in vitro* capacitation and fertilization. Approximately 10⁶ motile spermatozoa per ml of medium are frequently used for inseminating human eggs although concentrations ranging from 0.5 to 2.0 × 10⁶ per ml have been employed. The duration of insemination, using washed spermatozoa, resulting in proven fertilization has ranged from 3 to 24 hours^{9,12}. It is not known whether the conditions and media currently being used are optimal for human *in vitro* fertilization as adequately controlled experiments and statistical evaluations have not yet been reported.

Ham's F10 medium has been used most successfully for completing syngamy and initiating cleavage in fertilized human eggs, and for supporting development of the embryo up to the blastocyst stage^{13,14}. This medium has been reported to be favourable for human preimplantation embryo growth at osmolalities ranging from 280 to 300 mOsmol per kg, and at pH levels from 7.2 to 7.4 in an atmosphere of 5% CO₂ + 5% O₂ + 90% N₂ at 37°C. As a general rule the medium is supplemented with either inactivated human serum, fetal calf serum or human cord serum, at levels ranging from 15 to 30 per cent. Other additions including antibiotics, glutamine, calcium, lactate, pyruvate and crystalline albumin may have been used yet these modifications are not reported in publications. Investigators tend to introduce changes in the culture media and in the procedures they use. The exact culture media used, the supplements added and their final concentration, the final pH and osmolality of the insemination and culture medium, and the duration of

incubation before fresh medium (if any) was introduced into the culture system, have not yet been sufficiently detailed by Edwards *et al.*¹ to enable others working in this area to duplicate the laboratory procedures which have resulted in the four pregnancies following IVF and ET.

Human embryos developed *in vitro* and transferred into the uterus through the cervix have a very low chance of establishing a pregnancy that will continue. The high wastage of transferred embryos may be due to defects in the embryo, endometrium, corpus luteum or the transfer procedure. A cleaving embryo placed in the uterus may fail to develop into a normal blastocyst or it may carry a chromosomal abnormality which causes its

early death. The luteal endometrium may become asynchronous with the stage of development of the embryo, or non-receptive as a result of the absence of humoral or hormonal factors released by the cultured embryo. An inadequately functioning corpus luteum may result from damage sustained by the follicle during oocyte aspiration, or its function may not be sustained due to the absence of an embryonic signal. Poor transfer technique may lead to damage of the embryo, or endometrium, or to failed retention of the embryo in the uterus. It is proposed that a systematic evaluation of all these factors is required before the next significant advance in human IVF and ET will be achieved.

How heavy is a cold photon?

from N. Dombey

A welcome sign of the healthy state of theoretical physics today is the breakdown of the barriers which are insisted upon by grant-giving bodies and by most physics departments. Our research students in Britain, for example, must be strictly divided into astrophysicists (responsible to the SRC Astronomy, Space and Radio Board), nuclear or particle physicists (Nuclear Physics Board) and 'rest of physics' physicists (Science Board). But however useful (or useless) these divisions may be for experimentalists, they have been increasingly ignored by theorists over the last ten years or so.

The initial serious breakdown of these divisions came in the early 1970s as a result of Wilson's success with the application of renormalizable field theory to the classical problem of phase transitions; for the first time a sensible answer was provided to the question of why many critical indices are universal and independent of the substance undergoing the phase transition.

Many-body physics problems have for long been described in terms of thermodynamic, temperature-dependent Gibbs averages¹ rather than the zero-temperature vacuum expectation values of quantum field theory. It was also widely realised that a continuum limit of a crystal lattice system is applicable when the range of the excitations of interest is large compared with the lattice spacing as it is at a phase transition. But a systematic reduction of the different interactions of solid-state physics to the appropriate field theory came only with Wilson's work (see reference 2). Thus a Heisenberg ferromagnet can be reduced to the theory of an isovector scalar field ϕ with an interaction $|\phi|^2$ as for pions in particle physics. The critical exponents of the phase transitions can be calculated from the functions defined by the renormalization group equations of the field theory (for a

review see Kogut and Wilson²).

A phase transition represents a transition from one type of ordering or symmetry to another: in a ferromagnet, for example, all directions are equivalent above its transition temperature T_c (in the absence of an external magnetic field) so the system as a whole is rotationally invariant, but below T_c a ferromagnet will have a direction of magnetization which breaks this rotational symmetry. The interaction Hamiltonian which describes the behaviour of a ferromagnet for $T < T_c$ as well as $T > T_c$ is rotationally invariant — but the theory shows that the ground state of the system for $T < T_c$ is one with the spins preferentially pointing in the same direction and thus does not satisfy rotational invariance. This is an example of 'spontaneous symmetry breaking' and it is characterized by the ground state not possessing the full symmetry of the interaction describing the system. Such theories have been studied now for many years both in solid-state theory (condensed-state theory is a better term to include superconductivity and superfluidity) and in particle theory where the Higgs mechanism of spontaneous symmetry breaking is used in analogy with the theory of superconductivity to generate mass. The best known case of this is the spontaneously broken unified gauge theory $SU(2) \times U(1)$ of weak and electromagnetic interactions (*Nature* **282**, 131; 1979) where the mass of the gauge particles $W^\pm$, Z^0 is generated in this way although in the underlying unbroken gauge theory they would have zero mass.

Thus a class of field theories with similar properties has been studied during the past ten years in both particle theory and condensed-state theory. The linkage with astrophysics and cosmology has more recently paralleled these studies, motivated in particular by the asymmetry of baryons (protons and neutrons) in the universe. According to the generally accepted big-bang theory the universe has been

1. Edwards, R.G., Steptoe, P.C. and Purdy, J.M. *Br. J. Obstet. Gyn.* **87**, 737 (1980).
2. Lopata, A., Johnston, I.W., Hoult, I.J. and Speris, A.L. *Fertil. Steril.* **33**, 117 (1980).
3. Steptoe, P.C., Edwards, R.G. and Purdy, J.M. *Br. J. Obstet. Gyn.* **87**, 757 (1980).
4. de Crespigny, L.Ch., O'Helihi, C. and Robinson, H.P. (in preparation).
5. Fraser, L.R. and Dandekar, P.V. *J. exp. Zool.* **184**, 303 (1973).
6. Bavister, B.D. *J. Reprod. Fertil.* **13**, 544 (1969).
7. Edwards, R.G., Bavister, B.D. and Steptoe, P.C. *Nature* **221**, 632 (1969).
8. Soupart, P. and Strong, P.A. *Fertil. Steril.* **25**, 11 (1974).
9. Lopata, A., Brown, J.B., Leeton, J.F., Talbot, J.M. and Wood, C. *Fertil. Steril.* **30**, 27 (1978).
10. McMaster, R., Yanagimachi, R. and Lopata, A. *Biol. Reprod.* **19**, 212 (1978).
11. Hoppe, P.C. and Pitts, S. *Biol. Reprod.* **8**, 420 (1973).
12. Lopata, A., McMaster, R., McBain, J.C. and Johnston, W.I.H. *J. Reprod. Fertil.* **52**, 339 (1978).
13. Steptoe, P.C., Edwards, R.G. and Purdy, J.M. *Nature* **229**, 132 (1971).
14. Lopata, A. and Kohlman, D. *Abstracts of Australian Obstetrical and Gynaecological Research Society, Armidale* (1980).

N. Dombey is in the Department of Physics, University of Sussex.

expanding, and thereby cooling, from a state with infinite temperature and density, and the matter in the universe was created in the early stages of this process. Yet in spite of the law of baryon conservation, the universe contains large numbers of protons and neutrons, not anti-protons and anti-neutrons. Sakharov³, in a pioneering and far-sighted paper in 1967, tried to understand how this was possible. He assumed that baryon conservation was not absolute; so anti-protons could decay into electrons or negative muons (but then protons could also decay into e^+ or μ^+). As it was known, however, that CP was violated by a small amount (C is charge conjugation which turns a particle into its anti-particle; P is parity inversion), the rates for proton and anti-proton decay could differ and hence a proton and electron predominance could be built up.

This was viewed as a wild speculation for many years. With the apparent successes of the unified $SU(2) \times U(1)$ gauge theory of weak and electromagnetic interactions and of quantum chromodynamics (QCD — the $SU(3)_c$ gauge theory of strong interactions *Nature* 279, 479; 1979), however, many theorists have in recent years considered grand unified gauge theories of which the simplest is the $SU(5)$ theory of Georgi and Glashow⁴ (see also this issue of *Nature* p. 649). This is a theory unifying strong, electromagnetic and weak interactions which is broken by the Higgs mechanism first into the unbroken $SU(3)_c$ of strong interactions and the $SU(2) \times U(1)$ of unified weak/electromagnetic interactions, which itself is then broken by the Higgs mechanism in the usual way. A common feature of the grand unified theories is that they contain gauge particles which couple baryons to leptons so that baryon conservation is no longer an exact law.

In order to estimate the size of baryon non-conservation, which would be reflected in a finite rather than infinite proton lifetime, it is necessary to calculate the masses of the gauge particles which mediate these processes. This can be done in the same way as for $W^\pm$ and Z^0 in $SU(2) \times U(1)$: the characteristic mass here is about 100 GeV which is the energy where weak interactions and electromagnetic interactions would have the same strength. Similarly the energy at which electromagnetic interactions become strong is about 10^{15} GeV which therefore is the mass scale of these new particles. Thus the prediction of these theories is that baryons can decay into leptons but, as they would do so only by exchanging a superheavy gauge particle, this happens very rarely.

So there is a hierarchy of gauge theories

$$\begin{array}{lcl} SU(5) & \xrightarrow{10^{15} \text{ GeV}} & SU(3)_c \times SU(2) \times U(1) \\ & \xrightarrow{100 \text{ GeV}} & SU(3)_c \times U(1) \end{array}$$

where the fully symmetric $SU(5)$ theory is only realised at energies large compared with 10^{15} GeV so that the mass of all the gauge particles can be neglected; at

energies smaller than 10^{15} GeV but much larger than 100 GeV the mass of $W^\pm$ and Z^0 can be neglected so that weak and electromagnetic interactions are fully unified; and in our normal low-energy region of laboratory experiments (< 100 GeV) only QCD ($SU(3)_c$) and QED ($U(1)$) remain unbroken symmetries.

This is the present understanding of particle physicists who are eagerly waiting for an example of proton decay. Nevertheless, this hierarchy of gauge theories has immediate cosmological implications for the reasons advanced by Sakharov. We can consider the particles of these gauge theories as interacting with a background heat bath of temperature T . The appropriate theory would then be applicable to a very hot, very dense gas of particles interacting with the couplings of the gauge theory, provided that this gas had time to come to thermal equilibrium. So now Sakharov's speculation can be given flesh: at temperatures $T > 10^{28}$ K corresponding to energies above 10^{15} GeV this gas of radiation and particles will interact through a grand unified theory such as $SU(5)$ with its 24 zero mass gauge particles, and baryons will be created which can then decay into leptons. As the temperature decreases, there is now the possibility of a phase transition of the gas into a new phase with lower symmetry: this has been studied and can indeed happen (see Linde⁵ for a general review of the field). In particular, the present baryon-asymmetric state of the universe could be reached in this way, for a baryon asymmetry may have been produced in the early universe which was then 'frozen' in ever since⁶. So the hierarchy above can now be viewed schematically as

$$\begin{array}{lcl} SU(5) & \xrightarrow[10^{15} \text{ K}]{10^{28} \text{ K}} & SU(3)_c \times SU(2) \times U(1) \\ & & SU(3)_c \times U(1) \end{array}$$

where the dense gas at temperatures above 10^{28} K has the highest symmetry and various phase transitions occur to states of lower symmetry until we reach the present universe with its characteristic cosmic microwave temperature of 2.7 K at which only $SU(3)_c$ and $U(1)$ are good symmetries and of the original 24 gauge particles, only the 8 gluons of QCD and the photon of QED remain massless.

This brings us to the letter of Primack and Sher (p. 680, this issue of *Nature*): why they ask, should this hierarchy of broken gauge theories stop at $SU(3)_c \times U(1)$ and $T = 10^{15}$ K? Suppose at low enough

temperatures there is a further phase transition into a state of lower symmetry. This could imply "that the $SU(3)$ gauge symmetry of the strong interactions, quantum chromodynamics, is broken at low temperature. Is it possible that the reason Fairbank and collaborators have been the only group to detect fractionally charged particles — quarks? — is that theirs is the only low temperature search?" This suggestion is due to Deshpande.

Alternatively, or additionally, there is an even more intriguing possibility. Perhaps the $U(1)$ symmetry of QED is spontaneously broken. This would imply that its gauge particle, the photon, would become massive at a low enough temperature, or more exactly, that a photon gas in equilibrium with a heat bath would undergo a phase transition at $T = T_c$ so that at temperatures below T_c each photon in the gas would have a non-zero mass $\mu(T)$. There would be various manifestations of a photon mass μ . Following the review by Goldhaber and Nieto⁷, one can list four different effects and types of experiment:

- (i) the velocity of light would vary with its energy (that is, frequency)
- (ii) the electrostatic field would arise from a potential of Yukawa form ($e^{-\mu r}$) rather than the $1/r$ Coulomb potential. Here $\hbar = c = 1$.
- (iii) Ampère's law is violated and the corresponding Maxwell equation becomes $\nabla \times H = J - \mu^2 A$ where A is the vector potential; this also leads to an exponential fall-off proportional to $e^{-\mu r}$ of H for a magnetic dipole such as the earth, and
- (iv) the dispersion relation between frequency and wave-vector k becomes $\omega^2 = k^2 + \mu^2$ in place of $\omega^2 = k^2$. This will give a change in arrival time of radio signals from a pulsar dependent on frequency. This can be accurately measured.

Each of these effects gives a bound on μ with the best (precise measurements of the Earth's magnetic field) giving $\mu \leq 3 \times 10^{-15}$ eV, and the others giving upper limits between 10^{-10} and 10^{-14} eV.

Primack and Sher note that all these experiments are performed either at normal temperatures ($\sim 10^2$ K), or in the case of astronomical measurements, at 2.7 K. Thus they suggest that the photon mass be measured at temperatures much below 2.7 K. One such experiment is shortly to be performed at Sussex University by Clark and his collaborators. It will consist of looking for a frequency shift of a resonant mode in a superconducting cavity as the temperature is reduced in steps from an initial temperature of 4 K, where the photon is certainly massless, to a temperature of 0.05 K. It should be possible to measure the frequency shift (from (iv) above) arising from a phase transition giving the photon mass μ to one part in 10^8 ; this should give a mass limit at these temperatures of about 10^{-9} eV. □

1. Abrikosov, A.A., Gorkov, L.P. & Dzyaloshinski, L.E. *Methods of Quantum Field Theory in Statistical Physics* (1963) (translated and edited by R. Silverman, Dover, 1975).
2. Kogut, J. & Wilson, K. *Phys. Rep.* 12C, 75 (1974).
3. Sakharov, A.D. *Pis'ma Zh. Eksp. Teor. Fiz.* 5, 32 (1967).
4. Georgi, H. & Glashow, S.L. *Phys. Rev. Lett.* 32, 438 (1974).
5. Linde, A.D. *Rep. Prog. Phys.* 42, 389 (1979).
6. Yoshimura, S. & Susskind, L. *Phys. Rev. D* 18, 4500 (1978); Toussaint, D., Treiman, S., Wilczek, F. & Zee, A. *Phys. Rev. D* 19, 1036 (1979); Weinberg, S. *Phys. Rev. Lett.* 42, 850 (1979); Ellis, J., Guillard, M. & Nanopoulos, D. *Phys. Lett.* 80B, 360 (1978).
7. Goldhaber, A.S. & Nieto, M.M. *Rev. Mod. Phys.* 43, 277 (1971).



Selfish DNA

In two review articles in the 17 April issue Doolittle and Sapienza (p. 601) and Orgel and Crick (p. 604) separately suggested that much of the DNA in the genome of higher organisms could be described as 'selfish'. They argued that such DNA has no appreciable phenotypic effect and functions only to ensure its own self-preservation within the genome. This view point stimulated a great deal of comment, some of which was published in the issue of 26 June (p. 617). Now the original authors have joined up with one of their critics and reassessed their ideas in the two articles below. A further comment is added by H. K. Jain.

from L.E. Orgel, F.H.C. Crick and C. Sapienza

DIFFICULTIES have been caused by the words 'selfish', 'junk', 'specific' and 'phenotype' that were used in the two reviews of selfish DNA^{1,2}.

Many people dislike the term 'selfish DNA' and a more acceptable alternative might be 'parasitic DNA'. The word 'parasitic' does not imply that the DNA can move between individuals, though certain viral DNAs might do this. It does imply that such DNA can usually move between different chromosomes in the same cell.

The word 'junk' also seems to arouse strong feelings. The idea behind it can be clarified by considering what is meant by 'specific'. We consider a sequence highly specific if the change of any one of its bases almost always has a considerable effect on the organism. An example would be the recognition site for a physiologically relevant restriction enzyme. At the other extreme are sequences whose deletion or extensive alteration would produce a negligible effect. Such sequences could reasonably be called junk. However, there is probably a continuum between these two extremes, including fairly specific sequences, where the alteration of most bases will produce some effect (many sequences coding for protein, and the different signals for starting and stopping transcription are likely to be of this type) and sequences whose deletion or extensive alteration will usually produce a small effect, such as a change in the local rate of recombination. In some cases close similarity of two sequences may be important, while the base sequences themselves may matter hardly at all — for example, within the introns of two neighbouring versions of a gene. The word 'junk' is perhaps too broad to cover all those cases for which the effect of sequence on the phenotype of an organism is small or zero. We hope a more precise terminology will evolve as the facts become better known.

The word 'phenotype' has also caused difficulties in spite of Doolittle and Sapienza's careful use of 'organismal

phenotype' to make their meaning clear. We obviously need two words: one to refer to the phenotype of the organism and the other to apply solely to the 'phenotype' of the parasitic DNA, a distinction we would certainly make in the case of a true parasite. For the former we would suggest 'organismal phenotype' and for the latter, following Cavalier-Smith³, 'intragenomic phenotype', but we would allow the word 'phenotype' alone to be used when the context makes the meaning clear.

In our original definition we said that selfish DNA had two distinct properties: (1) It arises when a DNA sequence spreads by forming additional copies of itself within the genome. (2) It makes no specific contribution to the phenotype. By 'phenotype' we meant organismal phenotype. We intended 'specific' to be understood as 'highly specific' or 'fairly specific' in the discussion above. However it has been pointed out to us by R. Pritchard⁴ that 'no... contribution' is unnecessarily strict. It would have been more useful to include also DNA which made a small contribution to the organismal phenotype, either positive or negative. An example of the latter might be a viral DNA which became part of the genome.

There is obviously a continuum of possible selective advantages (positive or negative) to the organism. We had excluded from our definition of selfish DNA those cases where the selective advantage is very high. To decide whether a repeated sequence is parasitic or not, one must determine whether the presence of the repeated sequence in the population is mainly due to the efficiency with which the sequence spreads intragenomically or mainly due to the reproductive success of those individuals in the population who possess repeated copies of the sequence. Only in the former case do we consider it useful to use the term selfish or parasitic DNA, as opposed to useful or symbiotic DNA — the borderline between the two may not be sharp.

In considering the spread of parasitic DNA one should not underestimate the power of natural selection. For example, if a particular transposon was inserted at

random, it would run the risk of inactivating many genes and thus be selected against. A transposon which usually inserted at sites between genes would be at a selective advantage. Sites very near essential genes (as pointed out by Bruce Grant⁵) may be harder to delete than those in the middle of long stretches of junk and so parasitic DNA in the former positions is likely to survive longer. Effects of this type would lead to the selection of selfish DNA sequences that inserted preferentially at special sites in the genome.

Competing theories differ in their analysis of the factors determining the amount of non-specific DNA and of the way in which it comes into existence. Although we cannot at present decide on the quantitative contribution of the different types of non-specific DNA to the genome, it is still helpful to classify the various theories.

We proposed² that the amount of non-specific DNA present in a given genome is often determined by the balance between the intragenomic spreading of selfish sequences and phenotypic selection against excess DNA — the weaker the phenotypic selection against non-specific DNA the larger the DNA content of the genome. In another group of theories it is proposed that there is an optimal DNA content for each organism, which may be substantially greater than the amount of specific DNA that is needed to define the phenotype. The amount of non-specific DNA is then principally determined by the difference between the optimal DNA content and the essential content of specific DNA. The theories are not mutually exclusive, but differ substantially in emphasis in their explanation of C-values.

Cavalier-Smith's proposal^{3,6} is an interesting example of an 'optimal DNA content' theory. One of his ideas, which we misinterpreted in our previous paper², is that in large cells, particularly in oocytes, the transport of messenger RNA across the nuclear membrane may become a limiting factor and that the only way to increase the

The "selfish DNA" design at the top of the page was created by Linda Angeloff-Sapienza of Halifax, Canada and originally appeared on the cover of the issue of 17 April, 1980.

L.E. Orgel and F.H.C. Crick are at The Salk Institute, San Diego, California and C. Sapienza is in the Department of Biochemistry, Dalhousie University, Halifax, Canada.

rate of transport is by increasing the number of nuclear pores by extending the surface of the membrane. If the area of nuclear membrane is determined by the DNA content of the nucleus, it follows that selection for a larger cell must lead to an increase in the DNA content of the genome. Thus, rather surprisingly, extra non-specific DNA is selected for because it allows such a cell to grow *faster*. While we do not question the logic of the argument, given the various assumptions, we do not find all the assumptions particularly plausible. It may be that there is sometimes selection for increased cell volume and increased nuclear volume. In cells so selected, non-specific DNA can accumulate. Whether it does so because large cells with large nuclei require such accumulation, or because they simply permit it remains to be seen. We feel that more experimental work is needed to unravel the complexities of the situation. In particular, we should like to know in which stages and in which organisms the surface of the nuclear membrane is saturated with nuclear pores.

Cavalier-Smith³ also cites the widely different DNA contents of germ cells and somatic cells in some invertebrates as evidence against the selfish DNA hypothesis. However these observations can also be explained in terms of the selfish DNA theory. Such DNA 'needs' only to remain in the germ line to function parasitically. On the other hand, organismal selection might sometimes be stronger against surplus DNA in the soma than in the germ line. Thus representation in the germ line but not in the soma may sometimes be an optimal strategy for parasitic DNA. As for B chromosomes, in many cases the evidence appears to us to give some support to the idea (originally proposed by Östergren⁷ in 1945) that they are largely parasitic, but there is certainly evidence that they sometimes have phenotypic effects which may possibly be useful^{8,9}.

Smith¹⁰ has pointed out that the DNA of vertebrates usually has about 42 per cent GC whereas the GC content of invertebrate and prokaryotic DNA varies over a much wider range. The theory of parasitic DNA has rather little to say on this point. There are many factors which might affect the GC content of an organism's DNA. If much of the parasitic DNA has descended rather recently from insertion elements which themselves originally coded for proteins, then it would not be surprising if their present GC content were similar to

that of genes which still code for protein. This may, perhaps, explain the constancy of GC in vertebrates.

As for our own ideas, we now feel that there may perhaps be reasons why too *little* DNA can in some cases produce a selective disadvantage. For example, Zuckerkandl¹¹ has suggested that there may be a minimum size for a 'domain' necessary for stability of the chromatin in the folded state. Thus a domain containing only a few genes might benefit from having some non-specific DNA as 'padding'. This would mean that there is indeed an optimal amount for total DNA.

In our original paper² we feel that we did not put enough emphasis on the distinction between sequences which are repeated, exactly or nearly exactly, in many tandem repetitions and sequences which are more widely dispersed over the chromosomes and which occur in only one or a few copies in any one place. It seems plausible that these two types of sequence evolved different mechanisms. It is possible that the mechanisms generating the tandemly repeated type are usually more 'ignorant' (in Dover's sense¹²) than the more dispersed type. If the latter have any specific function it is likely to be that of the control, at one level or another, of gene

expression, whereas the tandemly repeated type seem more likely to influence chromosome mechanics.

One possibility to which we feel we should have given more weight is that of 'dead genes', also called 'pseudogenes'^{13,14}; that is, sequences which can no longer code for a protein (or a structural RNA) but which appear to have descended from a sequence that did. Whether these conform to our definition of parasitic DNA remains to be seen, but we suspect this is unlikely, since they usually exist in only a single copy, or as multiple tandem copies in only one place.

In our recent experience most people will agree, after discussion, that ignorant DNA, parasitic DNA, symbiotic DNA (that is, parasitic DNA which has become useful to the organism) and 'dead' DNA of one sort or another are all likely to be present in the chromosomes of higher organisms. Where people differ is in their estimates of the relative amounts. We feel that this can only be decided by experiment. We expect that due to the recent advances in genetic engineering and related techniques much sequence information will accrue in the near future. This should help to decide between the different alternatives. □

Modes of genome evolution

from Gabriel Dover and W. Ford Doolittle

OUR original articles^{1,3} presented antagonistic positions and perhaps obscured many areas of agreement. In essence, we are approaching similar phenomena from different perspectives and it might be useful to clarify where our views agree and differ and to indicate the sorts of evidence which would allow one to decide whether the origin of non-coding DNA in eukaryotic cells is more precisely viewed as 'ignorant'³ or 'selfish'^{1,2}.

The eukaryotic genome is constantly in flux, and this plasticity results from a variety of known and as yet mysterious mechanisms which amplify and disperse segments of DNA throughout a set of chromosomes⁴⁻⁷. Replication and recombination are complex processes requiring many enzymes that have evolved by natural selection. What we both wish to stress is that, in establishing these processes, evolution has inadvertently endowed the genome with built-in mechanisms for irregular and recurrent random sequence rearrangements and

created an environment in which elements capable (to varying extents) of promoting their own amplification and dispersion will inevitably arise^{1,2,7}. We acknowledge that there is evidence which suggests that some proportion of genome rearrangements may have effects on the biology of an organism (for instance on chromosome behaviour, on nuclear RNA processing, on cell-cycle times, on recombination frequencies and on gene expression; see refs 1-3, 8). Where this is the case, the change in frequency of a sequence rearrangement and the behaviour of elements which promote their own amplification and dispersion will of course depend on the effects they have on fitness. Hence, the net accumulation of these particular families of sequences reflects a balance between the intrinsic rate of accumulation (intra-genomically) and the effect of natural selection on phenotypic differences.

We agree that the amplification and dispersion of segments may occur either at random (sequence-independent or 'ignorant') or with preference for certain sequences (sequence-dependent or 'selfish'). Although both 'ignorant' and 'selfish' are unfortunate terms, they should be understood in the spirit in which they are

1. Doolittle, W. F. & Sapienza, C. *Nature* **284**, 601 (1980).
2. Orgel, L. E. & Crick, F. H. C. *Nature* **284**, 604 (1980).
3. Cavalier-Smith, T. *Nature* **285**, 617 (1980).
4. Pritchard, R. (personal communication).
5. Grant, B. (personal communication).
6. Cavalier-Smith, T. *J. Cell. Sci.* **34**, 247 (1978).
7. Östergren, G. *Bot. Notiser* **2**, 157 (1945).
8. Jones, R. N. *Int. Rev. Cytol.* **40**, 1 (1975).
9. Ames, A. & Dover, G. *Chromosoma* (in the press).
10. Smith, T. F. *Nature* **285**, 620 (1980).
11. Zuckerkandl, E. (personal communication).
12. Dover, G. *Nature* **285**, 618 (1980).
13. Loomis, W. *Dev. Biol.* **30**, F3-F4 (1973).
14. See, for example, Proudfoot, N. *Nature* **286**, 840 (1980).

Gabriel Dover is in the Department of Genetics, University of Cambridge, UK, and W. Ford Doolittle is in the Department of Biochemistry, Dalhousie University, Canada.

used and defined: sequence-independent and sequence-dependent processes of amplification and dispersion respectively.

The 'selfish process' produces DNAs which are preferentially chosen either by virtue of their nucleotide sequence or by virtue of the fact that they may code for gene products for their own amplification and dispersion. Mobile elements such as bacterial insertion sequences and transposons, the 'Ty-1-like' elements of yeast, the 'copia-like' elements of *Drosophila*, and vertebrate retroviruses, which show surprising similarities in structure and (perhaps) dispersal mechanisms⁸ are almost certainly the self-perpetuating products of this sort of process. An alternative term 'self-selection' might usefully describe the process of accumulation of these sequences. Interestingly, there may be an element of self-selection in the process of accumulation and dispersion of some sequence-independent segments. Extensive sharing of similar sequence patterns of repetitive DNAs between chromosomes (see refs 9 and 10), and occasionally between species^{11,12}, suggests that an arbitrarily accumulated sequence arrangement preferentially enjoys further rounds of amplification and dispersion¹⁰. Simulation of unequal recombination by computer appears to show a degree of self-perpetuation of sequences initially chosen at random¹³ and also that amplification and dispersion can be part and parcel of the same recombinational irregularity¹⁴.

A process of recurrent amplification can explain the frequent observation of a greater within-species than between-species sequence homogeneity of shared families^{6,9,12}. Similarly, polymorphisms and variations in sequence patterns of ribosomal genes¹⁵, histone genes¹⁶ and some non-coding families¹⁷⁻¹⁹ in several diverse organisms can be interpreted as the most recent, and hence localized, amplifications of these sequences.

We wish to emphasize, however, that 'ignorant' and 'selfish' self-perpetuation

are terms that uniquely apply to these particular DNA processes and cannot be used, meaningfully, to describe changes in frequencies of other elements that are totally dependent on the natural selection of phenotypes. 'Selfish genes'²⁰ and 'replicator selection'²¹ are not synonymous terms and the evolutionary processes to which they allude are unknown. A very limited accumulation of supernumerary B chromosomes is the only other process that can be described as self-accumulation, often the result of meiotic and mitotic non-disjunction. We doubt if this term can be used to describe the mis-named 'meiotic drive' mechanisms of regular A chromosomes in the rare instances where this occurs. For example, in the case of segregation distortion (SD) in *Drosophila melanogaster*, the preferential recovery of the SD chromosome is not so much due to its accumulation *per se* but is the outcome of dysgenesis of cells carrying the non-SD homologue²². This, and other cases, are analogous to the relative changes in frequency of alleles at a gene locus that are the outcome of natural selection; and 'selfish' and 'replicator selection' are misleading descriptions of this process of differential accumulation of alleles and chromosomes.

We do not agree upon the relative contributions of randomly accumulated and preferentially accumulated DNAs to the evolution of eukaryote genomes. It is clear that much of the non-coding DNA of yeast could be of the sort one could call 'selfish'^{4,8} whilst non-coding elements of the *D. melanogaster* genome are made up of varying proportions of essentially 'ignorant' repetitive DNA (for example the satellite DNAs²³) and 'selfish' transposable DNAs (for example the 'copia-like' elements⁵). The scrambled arrangements of repetitive elements in *D. melanogaster* observed by Wensink and co-workers²⁴

could reflect either sequence-independent shuffling processes or sequence-dependent insertion of transposable elements, similar to *copia*, at adjacent sites. The available data on repetitive DNA families of species of sea urchins, Gramineae, rodents, primates and insects (see ref.3) do not permit a clear assessment of the mechanisms that gave rise to them, for no direct sequence data are available. The multiple-copy 'Alu I' family of mammalian genomes may have a cellular function^{25,26}, but it is not impossible that these are the descendants of a family of transposable elements²⁷.

Finally, we suggest that the accumulation of sequences might be affected by constraints on the mechanisms themselves. For example, sequence-dependent amplification and dispersion will favour sequences that accurately contain the required sequence for transposition and that have not drifted too far into unacceptable divergent sequences. Similarly, sequence-independent mechanisms may be constrained to particular lengths of sequences¹⁰. Such constraints impose a type of selection on sequences, not necessarily as a result of their phenotypic effects but more as a consequence of the molecular mechanisms of replication and recombination.

We do not know what proportions of the repetitive DNAs of 'higher' organisms are amplified and dispersed by either sequence-dependent or sequence-independent mechanisms. Similarly, we do not know to what extent each of these mechanisms is constrained nor do we know the extent to which the frequencies of sequence patterns are an outcome of their possible effects on individual fitness. It is clear, however, that there are several modes of sequence rearrangement within rapidly evolving genomes. The problem now, as with most scientific debates, is one of quantification.

Incidental DNA

from H.K. Jain

THERE can be two ways of looking at the extraordinarily large amounts of DNA in the eukaryotic genome. One is to keep looking for functions (mostly of a non-conventional nature, for example, determining nucleotypic effects¹), and the other is to realise that lack of function is precisely what one would expect from a large part of the excess DNA. After all, the most important attributes one would expect of any kind of genetic material in the context of its fundamental role in evolution would be (a) mutability and recombination ability to create variation and (b)

mechanisms which would help to conserve a large part of this variation in spite of continuous selection, which leads to rejection of a great deal of newly-produced variability. Mechanisms of this kind are already known to exist. Heterozygote advantage is perhaps the most important of them. Ford² has argued on the basis of Fisher's³ theory that heterozygote advantage has evolved because natural selection favours it as a major instrument for the preservation of variability and for keeping the population polymorphic.

Based on similar considerations, one would expect that the property of DNA to show mutability not only in the sense of classical gene mutations and chromosomal

1. Doolittle, W.F. & Sapienza, C. *Nature* **284**, 601 (1980).
2. Orgel, L.E. & Crick, F.H.C. *Nature* **284**, 604 (1980).
3. Dover, G.A. *Nature* **285**, 618 (1980).
4. Cameron, J.R. *et al.* *Cell* **16**, 739 (1979).
5. Potter, S.S. *et al.* *Cell* **17**, 424 (1979).
6. Flavell, R.B. *et al.* in *Genome Organization and Expression in Plants* (Plenum, 1980).
7. Dover, G.A. *Chromosomes Today* **6**, 105 (1977).
8. *Cold Spring Harb. Symp. quant. Biol.* **45** (1980).
9. Dover, G.A. *Nature* **272**, 123 (1978).
10. Brown, S.D.M. & Dover, G.A. *Nucleic Acids Res.* **8**, 781 (1979).
11. Brown, S.D.M. & Dover, G.A. *Nature* **285**, 47 (1980).
12. Donehower, L. & Gillespie, D. *J. molec. Biol.* **134**, 805 (1979).
13. Smith, G.P. *Science* **191**, 528 (1976).
14. Smith, T., Brown, S.D.M. & Dover, G.A. (unpublished results).
15. Wellauer, P.K. *et al.* *J. molec. Biol.* **105**, 487 (1976).
16. Cohn, R.H. & Kedes, L.J. *Cell* **18**, 855 (1979).
17. Cooke, H.J. & Hindley, J. *Nucleic Acids Res.* **6**, 3177 (1979).
18. Christie, N.T. & Skinner, D.M. *Proc. natn. Acad. Sci. U.S.A.* **77**, 2786 (1980).
19. Donehower, L. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **77**, 2129 (1980).
20. Dawkins, R. *The Selfish Gene* (Oxford Univ. Press, 1976).
21. Dawkins, R.A. *Tierpsychology* **47**, 61 (1978).
22. Crow, J.F. *Scientific American* **104**, 1 February (1979).
23. Brutlag, E.E. *et al.* *Cold Spring Harb. Symp. quant. Biol.* **42**, 1137 (1977).
24. Wensink, P.C. *et al.* *Cell* **18**, 1231 (1977).
25. Jelinek, W.R. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **77**, 1398 (1980).
26. Rubin, C.M. *et al.* *Nature* **284**, 372 (1980).
27. Bell, G.I. *et al.* *Nucleic Acids Res.* **8**, 4091 (1980).

H.K. Jain is in the Indian Agricultural Research Institute, New Delhi.

rearrangements (for example, of the *Oenothera* kind) but also out-of-step fast replication (or amplification) of some parts of it is one which natural selection would help to evolve and fix in a population. Such a property may not be ascribed to quirks of DNA replication as suggested by Dover⁴ but to selection favouring it. Natural selection is concerned not only with the existing variability but even more so with mechanisms which ensure its continued availability. If there is intragenomic selection leading to rapid build-up of some of the DNA sequences (the selfish DNA of Doolittle and Sapienza⁵ and Orgel and Crick⁶) we must treat this part of DNA as *incidental* to the fundamental process of mutability so vital for ensuring continued supply of raw material for the production of new genes. It does not follow that all of the DNA produced in this manner will, in fact, acquire a function. A large part of it (or even all of it) may not do so and may be eliminated only on an evolutionary time scale. Meanwhile, new DNA of the same and similar kind may continue to be produced so that at a given point of time there will always be large amounts of non-specific DNA. This fraction is best described as 'incidental' rather than 'selfish' DNA. We may call it incidental because it is a byproduct of the inherent property of mutability of the genome, a characteristic to which natural selection attaches great importance even if it leads to the production of repeated sequences and a wasteful deployment of energy. Viewed in

this light, non-functional DNA is very much a product of natural selection — a selection operating for mutability *per se*. Its relative abundance is probably a function of its nonfunctional nature for any other DNA which carries information of one kind or another would create genetic imbalance and would be quickly rejected.

It is important to recognise that economy has not been a major consideration in the process of evolution. Overproduction and rejection (both quick and prolonged) are not uncommon features of evolution; the excessive amount of DNA should be considered in the same context. Nature places considerable premium on playing safe so that it will not run short of raw material even if this means indiscriminate production leading to sequences which are destined to remain functionless. With our present understanding of the origin of life (far from complete) and its evolution through

several billion years, we must be prepared by now to accept some degree of confusion. After all, the evolutionary process is without any kind of central management, planning and coordination. It is totally dependent on variability arising through errors in the replication of DNA and its subsequent multiplication through recombination, which at times may involve genome fusion⁷. Whatever control mechanisms do exist are themselves a product of these errors which have been seized upon by natural selection to bring about some amount of order. A large number of mutant genes are now known which affect the meiotic process in various ways including chromosome pairing^{8,9}, chromosome condensation¹⁰, preferential segregation¹¹, amount of recombination^{12,13} and there are also, of course, the mutator genes^{14,15}, some of which could possibly be involved in the production of non-specific DNA.

The use of the term 'selfish' may be quite harmless if one does not try to see too much meaning in it. With our recollection of the reactions to Darwin's theory, it is too much to expect that attempts of this kind will not be made. We do not have to be squeamish when science demolishes some of our cherished beliefs. The truth pointed out by Darwin has, in the end, given a more meaningful direction to our social and cultural evolution. By the same token, we do not have to create unnecessary despair if the interesting concept of selfish DNA is misunderstood, as is likely to happen. □

1. Cavalier-Smith, T. *Nature* **285**, 617 (1980).
2. Ford, E.B. *Ecological Genetics* (Chapman and Hall, London, 1975).
3. Fisher, R.A. *The Genetical Theory of Natural Selection* (Oxford, 1930).
4. Dover, G. *Nature* **285**, 618 (1980).
5. Doolittle, W.F. & Sapienza, C. *Nature* **284**, 601 (1980).
6. Orgel, L.E. & Crick, F.H.C. *Nature* **284**, 604 (1980).
7. Schwesinger, M.D. *Bact. Rev.* **41**, 872 (1977).
8. Beadle, G.W. *Cytologia* **20**, 269 (1933).
9. Riley, R. *Heredity* **15**, 407 (1960).
10. Rhoades, M.M. & Dempsey, E. *Genetics* **54**, 505 (1966).
11. Rhoades, M.M. & Dempsey, E. *Genetics* **53**, 989 (1966).
12. Baker, B.S. *et al. A. Rev. Genet.* **10**, 53 (1976).
13. Rees, H. *Bot. Rev.* **27**, 288 (1961).
14. Cox, E.C. *A. Rev. Genet.* **10**, 135 (1976).
15. Yanofsky, C. *et al. Proc. natn. Acad. Sci. U.S.A.* **55**, 274 (1966).



100 years ago

NEW GUINEA
from L.M. D'Alberty

"The most perfect harmony seems to reign in families of the Papuan Mahoris, and rare indeed are cases of quarrel among members of one household. They live in communities, sometimes of more than a thousand inhabitants, in well-built villages, worthy to be called small towns, both for their order and cleanliness. They are under the rule of the chiefs or landowners. The chief is looked upon as father of the family. He is called Pacao, and his servant or subject is called Irine. From all I could learn, slavery does not exist, and the sale of human beings is unknown." After describing their daily avocations, amusements, dress, implements, and ornaments (a group of which are figured), he goes on: "Their natural disposition is gentle and placid. They like to spend their time in talking and games, in which men and women take an equal share. Playful and free of speech,



Durabi, a native of Kiwai Island, at the mouth of the Fly River

they nevertheless do not transgress the bounds of modesty, either in word or deed. Women and children are included in every conversation, and often take part in public discussions, which are usually held in the evening. Women are always respected, and in some villages they enjoy a certain supremacy, although the government of the house belongs to the husband. Labour may

be said to be fairly divided between the two sexes, and they are accustomed to work from their earliest childhood. . . . Will they be the happier for civilization? This is a difficult problem, and one which cannot be solved until the experiment has been made. For my part I do not doubt that these, more readily than any other savages whom I know, would answer to the call of a civilised nation which, stretching out a paternal hand, would lead them towards our civilisation! To insure success, however, they should be treated as friends, not as slaves; they should be cherished, not destroyed."

Unfortunately our attempts at civilising savages have as yet in every case failed. Are we still, notwithstanding all our wretched failures, to go on in the old way, and allow these interesting and now happy people to be first ruined morally by the teaching of the dregs of our Australian and Pacific traders, and then physically deteriorated by the forced introduction of a form of civilisation utterly unsuited to them? Cannot either philanthropy, or religion, or Government protect these people from all such external influences as have been proved to be unsuited to their condition and stage of development, while aiding them to work out for themselves an indigenous civilisation? Here is perhaps the last chance we have to preserve one remnant of the better class of savages from being crushed under the Juggernaut car of our high-pressure civilisation and mad struggle for wealth.

From *Nature* **23**, 23 December, 175, 1880.

REVIEW ARTICLE

Why unify?

Howard Georgi

Lyman Laboratory of Physics, Harvard University, Cambridge, Massachusetts 02138

Unified gauge theories (such as $SU(5)$) of particle interactions are built on the colour $SU(3)$ and $SU(2) \times U(1)$ gauge theories which apparently describe strong and weak and electromagnetic interactions at distances as small as 10^{-16} cm. In this article the classical reasons for going beyond $SU(3) \times SU(2) \times U(1)$ to a fully unified theory such as $SU(5)$ are reviewed, and a new reason formulated. A class of imaginary worlds similar to our own is considered and it is shown that unification is possible only in ours. This suggests that the low-energy interactions are unique in that they are constructed to make unification possible.

AS we enter the 1980s, elementary particle theorists have an unusual problem: we know (or think we know) too much. We confidently expect the standard $SU(3) \times SU(2) \times U(1)$ gauge theory of the strong, weak and electromagnetic interactions to describe particle physics accurately down to distances as small as 10^{-16} cm (ref. 1). The current status of particle physics has been reviewed by Mulvey². This article also briefly discusses the symmetry groups such as $SU(3)$, used to describe elementary particles in terms of quark multiplets.

The $SU(3)$ component of $SU(3) \times SU(2) \times U(1)$ is the colour $SU(3)$ theory of strong interactions or quantum chromodynamics (QCD) (the experimental and theoretical background which led to the development of QCD is reviewed in ref. 3). The eight $SU(3)$ gauge particles are the gluons. They couple to the three colours of the quarks through the eight traceless, hermitian 3×3 matrices. The effect of the colour $SU(3)$ interactions is to confine the coloured quarks into colourless hadrons (for which all the $SU(3)$ charges vanish).

The $SU(2) \times U(1)$ component is the Glashow–Weinberg–Salam theory of the weak and electromagnetic interactions. The four $SU(2) \times U(1)$ gauge particles are the $W^\pm$ and Z^0 (heavy because the $SU(2) \times U(1)$ symmetry is spontaneously broken) and the photon.

There are still important questions in the 'low-energy' (<100 GeV) domain of contemporary particle experiments. For example, given the colour $SU(3)$ theory, how, in detail, do quarks get confined? But the really fundamental questions seem to involve smaller distances and higher energies.

In 1973, it was suggested that the next step beyond $SU(3) \times SU(2) \times U(1)$ must be the unification of these three gauge groups as subgroups of $SU(5)$, so that all of the low-energy particle forces are components of a single, basic force⁴. For a large class of models, including $SU(5)$, it was shown that unification must take place at very short distances of the order of 10^{-29} cm (ref. 5).

But why should we go beyond the standard model at all? Here I review the obvious arguments for unification and suggest a new one.

Experimental evidence

It is appropriate to begin with a few words about the quality of experimental support for the standard $SU(3) \times SU(2) \times U(1)$ model. The evidence comes from many sources. There is much qualitative support for colour $SU(3)$ in hadron spectroscopy and for $SU(2) \times U(1)$ in the structure of weak decays. At a more quantitative level are lepton–lepton scattering and deep inelastic lepton–hadron scattering experiments. Summarizing this: although there is no single convincing datum (like $g-2$ of the electron for QED), there are many strong indications all pointing in the same direction. Taken together, the evidence for $SU(3) \times SU(2) \times U(1)$ is rather impressive.

Note that most of the quantitative support for $SU(3) \times SU(2) \times U(1)$ was obtained after the $SU(5)$ unification was proposed. Indeed, this is one of the main reasons that unification is taken more seriously now than when it was first suggested.

Family structure

Intertwined with all the evidence for $SU(3) \times SU(2) \times U(1)$ is the suggestive and mysterious family structure. Historically, the muon, its neutrino and the strange and charmed quarks were crucial in unravelling the structure of $SU(3) \times SU(2) \times U(1)$. But from a modern vantage point, they seem completely extraneous (and the τ lepton and b quark even more so). We would be happy with a world in which only the electron, its neutrino and the up and down quarks exist. In fact, in discussing unification, I will consider only the electron family, e^- , ν_e , and u and d.

$SU(5)$ unification

The left-handed (LH) fields transform as follows under $SU(3) \times SU(2) \times U(1)$

$$\begin{aligned} (\nu_e, e^-)_L &: (1, 2)_{-1/2} \\ e_L^+ &: (1, 1)_1 \\ (u, d)_L &: (3, 2)_{1/6} \\ u_L^c &: (\bar{3}, 1)_{-2/3} \\ d_L^c &: (\bar{3}, 1)_{1/3} \end{aligned} \quad (1)$$

where c indicates charge conjugate; the two numbers in parentheses indicate the $SU(3)$ and $SU(2)$ representations (in that order) while the subscript is the $U(1)$ quantum number (which is just the average electromagnetic charge of the multiplet); colour indices are suppressed.

In $SU(5)$, the $(\nu_e, e^-)_L$ and d_L^c are combined into a five-dimensional representation of $SU(5)$ which is conventionally called the $\bar{5}$. The corresponding right-handed (RH) fields, $(e^+, \nu_e^c)_R$ and d_R thus transform like a $\bar{5}$, the defining representation. The other LH fields transform like an irreducible 10-dimensional representation, the 10.

In this simple structure, the total electromagnetic charge in each multiplet is zero as it must be. Furthermore, any $SU(5)$ representation except a singlet must transform nontrivially under both the $SU(2)$ and $SU(3)$ subgroups, and these do. Finally, this system satisfies a nontrivial constraint, the absence of triangle anomalies, which must be satisfied for the theory to make sense^{6,7}. This constraint can be described as follows: write the gauge group generators as matrices acting on the left-handed fermion fields. Then the symmetrized trace of any product of three generators must vanish. In particular, for any three diagonal generators, the product of the eigenvalues summed

over all the LH fermion states must be zero. This constraint is satisfied for the generators of the $SU(3) \times SU(2) \times U(1)$ subgroup⁶⁻⁸ and also for the rest of the $SU(5)$ generators⁴.

Charge quantization

$SU(5)$ provides a very simple explanation for the experimental fact of charge quantization. The fundamental 5 describes leptons with charges 1 and 0 and quarks with charge $-1/3$. Because all $SU(5)$ representation can be built out of multiples of the 5, all charges are just sums of the fundamental charges and thus multiples of $1/3$. More than that there is a relation between colour $SU(3)$ and charge because the fractional charge $-1/3$ is carried only by the quarks in the 5. Thus triality zero states (and in particular colour singlet states) have integral charge. Triality one (two) states have charges $n - 1/3$ ($n + 1/3$) for integral n .

Given the specific charges in the 5, charge quantization is trivial because by inspection, all the charges in the 5 are commensurate. Actually the charge quantization theorem is deeper and it has a topological aspect. Whenever the unifying gauge group is compact, the charges have to come out commensurate⁴. This is also related to the fact that gauge theories based on simple groups have magnetic monopoles⁹ which in turn are related to charge quantization.

Coupling constant relations

The $SU(5)$ scheme predicts a relation between the gauge couplings of the three subgroups, so that for example, it predicts $\sin^2 \theta_w$ in terms of α and α_s (ref. 5). The quantitative prediction $\sin^2 \theta_w \approx 0.205$ is in reasonable agreement with experiment. But, the picture of coupling constant renormalization in which all are equal at very small distances and evolve differently at larger distances gives a satisfying qualitative explanation of the hierarchy of particle interactions. It explains why the strong interactions are strong.

Baryon number violations

The desire to explain charge quantization and to relate the different coupling constants motivated the search for unified theories from the beginning. Baryon number-violating interactions on the other hand were not sought after. Proton decay appeared unbidden in the $SU(5)$ scheme and elsewhere^{10,11}. But baryon number violation is now seen as an attractive artefact of unification. It yields the most plausible explanation yet proposed for the baryon number of the Universe¹². Furthermore, current experiments should see proton decay directly and give us an entirely new view on the particle physics world¹³. Pati and Salam¹⁰ have discussed baryon number violation in a partially unified theory. They have also considered¹¹ an alternative version of unification different from $SU(5)$ and derived independently and at about the same time. They assume that the colour symmetry is spontaneously broken and the quarks are 'liberated' with integral charge. The baryon-number violation proceeds by a completely different mechanism (depending on colour breakdown) from that in the $SU(5)$ model. In particular, it predicts baryon decay into final states with lepton number 3 rather than -1 .

Discussion

Besides these well known successes and predictions, the virtue of the $SU(5)$ unification is its extreme economy and simplicity. However, this simplicity sometimes makes people suspicious. Can unification really be as simple as ' $2 + 3 = 5$ '? I believe that this simplicity is deceptive. The theory is not so much simple as unique. I shall now try to prove the uniqueness of $SU(5)$ by discussing a class of imaginary worlds which cannot be unified.

First, I will argue that ' $2 + 2 \neq 4$ '. Consider a world like our own but with two colours instead of three. That is, the gauge group is $SU(2) \times SU(2) \times U(1)$ and the LH fields are as follows:

$$\begin{aligned} (\nu_e, e^-)_L &: (1, 2)_{-1/2} \\ e_L^+ &: (1, 1)_1 \end{aligned} \quad (2a)$$

as in equation (1); the quarks have charges $3/4$ and $-1/4$,

$$\begin{aligned} (u, d)_L &: (2, 2)_{1/4} \\ u_L^+ &: (2, 1)_{-3/4} \\ d_L^+ &: (2, 1)_{1/4} \end{aligned} \quad (2b)$$

The charges of the quarks are determined by the constraint of anomaly cancellation for the $SU(2) \times SU(2) \times U(1)$ theory.

Of course, this theory does not describe our world. We know that there are three colours and the quark charges are $2/3$ and $-1/3$. But as far as we know, equation (2) describes a possible alternative world. Thus why is our world described by equation (1) rather than equation (2)? The only important difference is that the two-colour theory cannot be unified. This can be seen easily. The 11 fields of equation (2) cannot be divided into two subsets, each of which has total charge zero and is nontrivial with respect to $SU(2) \times SU(2) \times U(1)$. Thus any unification must treat all 11 fields as a single irreducible representation. But then the group must be $SU(11)$ and the resulting theory does not satisfy the anomaly constraint.

Mathematically, we can describe the process of unification as follows: given a representation of a nonsimple group (such as $SU(3) \times SU(2) \times U(1)$) which satisfies the anomaly constraint, find an anomaly-free simple extension (such as $SU(5)$). As the example of $SU(2) \times SU(2) \times U(1)$ shows, this is not always possible—in fact, it is usually not possible.

It is straightforward to show that in the same sense it is impossible to unify a theory of standard leptons and quarks with more than three colours. That is, we can consider an infinite class of imaginary worlds with gauge group $SU(n) \times SU(2) \times U(1)$, $n > 3$, and LH fermions in a representation R (the quarks have charges $(1 \pm n)/2n$):

$$\begin{aligned} (\nu_e, e^-)_L &: (1, 2)_{-1/2} \\ e_L^+ &: (1, 1)_1 \\ (u, d)_L &: (n, 2)_{1/2n} \\ u_L^+ &: (\bar{n}, 1)_{-(n+1)/2n} \\ d_L^+ &: (\bar{n}, 1)_{(n-1)/2n} \end{aligned} \quad (3)$$

Again, the charge is chosen to cancel the $SU(n) \times SU(2) \times U(1)$ anomalies. Here, however, there is an argument against unification which does not involve the charge.

Assume that the unifying group is $SU(N)$, $N \geq n + 2$. This must be so because of the form of R . Only $SU(N)$, $0(4N + 2)$ and $E(6)$ have complex representations and the representations of $0(4N + 2)$ and $E(6)$ contain other fields besides the leptons and 'quarks' of equation (3). Now consider some conventionally normalized colour generator T_c . In the representation R of equation (3),

$$\text{tr}(T_c^2)_R = 2 \quad (4)$$

by explicit calculation. But in the N -dimensional representation, this same trace satisfies the inequality

$$\text{tr}(T_c^2)_N \geq 1/2 \quad (5)$$

because the N must contain at least one n of the $SU(n)$ subgroup. Now suppose that $R \neq N$ (and $R \neq \bar{N}$). We can derive a contradiction because the trace in any other representation is too large. It is easy to show that:

$$\frac{\text{tr}(T_c^2)_{R \neq N \text{ or } \bar{N}}}{\text{tr}(T_c^2)_N} \geq (N - 2) \geq 4 \quad (6)$$

The first inequality is pure group theory. It is clear from the form of R that it cannot be a direct sum of more than one N . The next smallest representation is the $N(N - 1)/2$ which saturates the first inequality. The second inequality follows because $N \geq$

$n + 2 \geq 6$ in the theories we are discussing. Comparing equations (4), (5) and (6) shows that $R \neq N$ (or $\bar{N}$) is impossible. Thus R must be the defining representation of $SU(N)$. This too is impossible because the theory has anomalies. The theory described by equation (3) for $n > 3$ has no anomaly-free simple extension.

Of course, there are anomaly-free simple theories which contain $SU(n) \times SU(2) \times U(1)$ for $n \neq 3$, but they must contain fermions beyond the usual quarks and leptons. $SU(3) \times SU(2) \times$

$U(1)$ is unique in that it permits unification of quarks and leptons.

This article is based on a talk given at the 1980 Cosener's House conference on Grand Unification. I thank the Rutherford Labs and in particular Chris Maxwell, Gordon Ringland and Mike Teper, and the staff of the Cosener's House for their hospitality and encouragement. I thank Tony Zee for a correction to this work. This work was partly supported by NSF grant PH77-22864.

1. De Rújula, A. *et al.* *Phys. Rev. D* **12**, 147 (1975).

2. Mulvey, J. *Nature* **278**, 403–409 (1979).

3. Marciano, W. & Pagels, H. *Nature* **79**, 479–483 (1979).

4. Georgi, H. & Glashow, S. L. *Phys. Rev. Lett.* **32**, 236 (1974).

5. Georgi, H., Quinn, H. R. & Weinberg, S. *Phys. Rev. Lett.* **33**, 451 (1974).

6. Bouchiat, C., Iliopoulos, J. & Meyer, Ph. *Phys. Lett.* **38B**, 519 (1972).

7. Gross, D. & Jackiw, *Phys. Rev. D* **6**, 477 (1972).

8. Pati, J. C. & Salam, A. *Phys. Rev. D* **8**, 1240 (1973).

9. Coleman, S. in *New Phenomena in Subnuclear Physics* (ed. Zichichi, A.) 297 (Plenum, New York, 1977).

10. Pati, J. C. & Salam, A. *Phys. Rev. Lett.* **31**, 661 (1973).

11. Pati, J. C. & Salam, A. *Phys. Rev. D* **10**, 275 (1974).

12. Nanopoulos, D. V. & Weinberg, S. *Phys. Rev. D* **20**, 2484 (1979).

13. Marshak, M., Winn, D., Steinberg, R. & Sulak, L. *First Workshop on Grand Unification*, (eds Frampton, P. H., Glashow, S. L. & Yildiz, A.) (Math Science Press, Brookline, 1980).

ARTICLES

Siderophile-enriched sediments from the Cretaceous–Tertiary boundary

Frank T. Kyte, Zhiming Zhou & John T. Wasson

Institute of Geophysics and Planetary Physics, University of California, Los Angeles, California 90024

Siderophilic element concentrations are high in sediments from the Cretaceous–Tertiary boundary. An extraterrestrial source is indicated. Concentrations are too high to be understood in terms of the impact of a chondritic asteroid. Either the projectile was a metal-sulphide core or the infalling material (probably weak cometary matter) was slowed down during atmospheric passage.

REPORTS by Alvarez *et al.*^{1,2} of anomalously high Ir concentrations in Cretaceous–Tertiary (CT) boundary sediments led us to analyse samples from its most enriched boundary, the fish clay from Stevns Klint, Denmark, and a Pacific deep-sea drill core which crossed the CT boundary. Alvarez *et al.* interpret the high Ir content as characteristic of an asteroid having a radius of 10 ± 4 km; they attribute the massive extinctions that marked the end of the Cretaceous to stratospheric opacity produced by impact-generated dust.

Because Ir is highly depleted in the Earth's crust relative to solar abundances, or the similar values in chondritic or iron meteorites, it is an ideal tracer for extraterrestrial sources. Other siderophiles (elements concentrated in an Fe–Ni phase if one is present) should also show high concentrations in the high Ir samples. Our procedure allows the rapid determination of several of these elements (Pd, Re, Os, Ir, Pt, Au) to compare their relative abundances in the CT boundary samples with those found in different groups of meteorites. Since starting this study there have been a similar but less comprehensive study of the Stevns Klint clay³, and data have been obtained on Ir, Os and several other elements in a boundary clay from Caravaca, Spain⁴.

Samples

Analyses were performed on 23 samples of the Stevns Klint fish clay and three samples from Deep Sea Drilling Project (DSDP) Leg 62, hole 465A. The Danish specimens 1 and 2 had thicknesses of 3.5 and 2.5 cm respectively, and were sliced parallel to the pronounced bedding to measure variations in composition with stratigraphic position; these yielded 9 and 6 samples, respectively. Unfortunately, we may not have sampled the actual base because siderophiles and sulphide-associated

elements are highest in our lowest samples, whereas Christensen *et al.*⁵ report a sharp decrease in concentrations of these elements in the basal, gray, layered marl. Eight specimens (numbers 4–11) which were dark coloured and clay-rich were also analysed; we infer that most of these originated in the lower portion of the section. Some have higher Ir concentrations than those determined in specimens 1 or 2.

Specimens 1 and 2 seem to contain three of the beds described by Christensen *et al.*⁵. In both specimens, sample 1 is the upper portion of a dark brown to grey marl, which must be either bed II or III. Both pairs of samples 2A and 2B are portions of a streaked grey marl corresponding to either bed III or IV. In sample 1, 3 this grades into the overlying lighter coloured marl which is certainly bed V. Here, the remaining samples are interbedded tan marls and fossiliferous chalk layers. The uncertainty in assigning the first two beds was due to: (1) the report⁵ that bed IV is intermittent, and found only in the deepest portions of the depositional basins; and (2) that Christensen *et al.*'s bed IV is thicker (3–5 cm) than our second unit (6 mm)⁵.

DSDP hole 465A is located in the Central Pacific at $33^\circ 49.23'N$ and $178^\circ 55.14'E$. Samples were taken from sections 3 and 4 in core 3. These sections primarily consist of a calcareous nannofossil–foraminifera ooze with the consistency of a thick paste; the Cretaceous–Tertiary boundary occurs within this ooze. The upper Cretaceous ooze is white and the lower Tertiary ooze is light tan. Unfortunately, drilling deformed the core and the boundary is smeared over at least 20 cm. Within this interval are several blebs of material containing pyrite (confirmed by our X-ray diffraction studies); these blebs may be fragments of a single 3-mm layer. We describe here the analysis of two samples containing pyrite (465A–1, 2), and one of the white ooze from below the boundary (465A–3).

Experimental

All samples were dried at 150 °C for 24 h, then ground to a fine powder. Aliquots of 200–400 mg were irradiated with flux monitors at UCLA in a neutron flux of $2 \times 10^{12} \text{ cm}^{-2} \text{ s}^{-1}$ for instrumental neutron activation analysis (INAA). Aliquots of 40–90 mg were leached with 1M HCl to determine CaCO_3 .

Separate aliquots of ~1 g of six Stevns Klint samples were similarly irradiated for radiochemical neutron activation analysis (RNAA) of the elements Ir, Os, Pt, Re, Au, Pd, and Ge. After dissolution samples were fused in Na_2O_2 , dissolved in HCl, and the noble siderophiles separated as a group on Sraffon resin⁶. Carrier recovery was determined by reactivation. Germanium was separated from the eluate by solvent extraction into CCl_4 and purified by distillation as GeCl_4 . The samples have high Br contents, and our determinations were seriously affected by neutron activated Br that was not easily eluted from the resin. The irradiated powders were therefore leached with H_2O before dissolution, reducing Br levels by a factor ≥ 10 . We doubt that this procedure removed a significant portion of the noble metals but confirmatory tests are in progress. Aliquots of Smithsonian standard Allende CV chondrite (split 3, pos. 18) were included in the RNAA runs as internal standards.

Results

Table 1 shows INAA data for 25 elements and two CaCO_3 determinations (the first based on loss on leaching with 1M HCl, the second based on the assumption that all Ca is present as CaCO_3). The good agreement (67% within 6%) by these independent methods suggest that the latter assumption is correct, and that negligible amounts of other carbonates are present; and also that only a minor fraction of the Ca is in the clay. Table 2 lists the RNAA data for seven elements in six samples of the Danish fish clay. These data confirm the high concentrations of Ir and other siderophiles in the lowest strata. Correction for CaCO_3 reduces the spread.

Our data on the Stevns Klint fish clay show good general agreement with those of Ganapathy; after allowing for anticipated sampling variations the only significant discrepancies seem to be our 3 times lower Re, 1.9 times higher Pt and 1.6 times lower Au values. Many of our data agree with those of Alvarez *et al.*¹ to within ~20%; we have compared their Stevns Klint samples with our basal samples 1, 1 and 2, 1 which have

similar Ir contents. The chief discrepancies are for a series of elements in which our values are lower by the listed factors: Co, 2.7; Ni, 2.1; Zn, 1.6; As, 2.4; Sb, 2.0. These elements are probably enhanced in strata with high pyrite content, thus the discrepancy can probably be attributed to sampling variations. Curiously, Se concentrations agree even though Se should also be enhanced in sulphide-rich samples.

Our data on DSDP hole 465A add another world location to CT boundaries having Ir concentrations far higher than typical values for pelagic clays⁷. Our Ir concentration of 16 ng per g is lower than those in the Stevns Klint and Caravaca boundary clays, but higher than that at Gubbio.

Elements in the Stevns Klint fish clay

Relative INAA concentrations of 10 elements are plotted in Fig. 1 as a function of height within Stevns Klint fish clay section 1; all concentrations have been recalculated on a CaCO_3 -free basis. Bulk CaCO_3 concentrations are also shown for reference. Only elements whose corrected concentrations either decrease or are constant within 10% are plotted (most lithophiles are not plotted).

The fish clay is finely layered and the variations observed in the section are probably associated with changes in the mineralogy. Thus the very high contents of As, Se and Sb in the lowest sample (Fig. 1) probably reflect higher concentrations of sulphides. But whereas Se levels remain about 6–10 times lower in all higher samples, As and Sb pass through a minimum 4 times lower than the basal sample then rise to final concentrations higher than that in the basal sample. If these high As and Sb concentrations in the higher levels also reflect a high content of sulphides, these were either of a sort that did not enhance Se, or the Se source had decreased greatly before their formation.

More relevant to the present discussion is the behaviour of Ir, which is initially high, decreases by a factor of 2.6 at a height of 15 mm, then gradually rises again to 0.6–0.8 times the basal value. Similar patterns are shown by Fe, Co, and Ni. The trends of Zn and Cr differ from those of Ni and Co chiefly in that the basal values are not appreciably elevated relative to the higher samples. Bromine concentrations are very high in the three lowest samples, then fall rapidly to a plateau 2–3 times lower. As we can leach 90% of this Br with H_2O , it is only weakly bound.

Table 1 Concentrations* of 25 elements in the Stevns Klint fish clay and in samples from sections 3 and 4, core 3 of DSDP hole 465A, Leg 62 from the Central Pacific

Height†	Na	K	Ca	Sc	Cr	Mn	Fe	Co	Ni	Zn	As	Se	Br	Sb	La	Ce	Nd	Sm	Eu	Tb	Dy	Yb	Lu	Ir	Th	CaCO_3	CaCO_3
1, 1	890	2.8	234	6.7	112	55	14.3	23.3	239	260	15.3	22.7	58.0	2.85	33	28	32	6.7	1.44	0.99	6.0	2.2	0.31	31.0	—	642	584
1, 2A	870	3.3	243	6.7	118	62	7.4	15.7	144	207	6.2	2.59	59.0	1.24	30	27	36	6.0	1.35	0.90	5.3	2.0	0.32	17.7	—	659	607
1, 2B	780	3.4	265	6.6	106	80	7.9	14.6	133	206	5.4	1.93	44.0	0.94	29	27	32	5.8	1.36	0.88	5.4	2.1	0.30	13.8	—	708	662
1, 3	890	3.4	309	5.1	56	105	5.6	10.4	82	144	3.2	1.50	15.0	0.55	27	22	26	5.4	1.25	0.84	4.8	1.9	0.27	7.0	—	720	772
1, 4	900	3.3	327	5.0	38	131	4.7	7.9	66	126	2.7	1.47	8.2	0.50	28	22	26	5.7	1.30	0.87	5.1	2.0	0.27	4.6	—	846	817
1, 5A	970	3.8	310	5.9	70	113	8.2	11.6	111	180	2.4	1.67	15.0	0.85	38	34	35	7.3	1.75	1.16	7.2	2.8	0.41	7.4	—	768	768
1, 5B	880	3.1	328	5.0	59	142	7.8	9.9	80	160	7.1	1.25	13.0	0.95	31	27	29	5.8	1.34	0.90	5.5	2.3	0.32	5.6	—	808	819
1, 6	1,450	1.7	370	4.2	38	160	5.5	6.4	60	109	5.6	1.36	9.1	0.88	45	39	40	8.8	2.00	1.34	8.6	3.2	0.41	5.8	—	889	924
1, 7	780	2.0	346	4.1	46	102	6.8	8.2	94	115	7.6	1.61	9.3	0.77	25	22	23	4.9	1.13	0.76	2.6	1.9	0.28	5.6	—	847	964
2, 1	680	2.7	224	7.6	140	42	17.0	31.0	298	297	21.6	16.70	69.0	3.47	32	30	35	6.4	1.40	0.94	5.7	2.0	0.30	36.0	—	594	559
2, 2A	710	3.5	242	6.9	131	62	9.1	17.8	176	220	9.1	2.71	57.0	1.61	27	25	29	5.4	1.23	0.83	4.8	1.8	0.28	19.6	—	667	604
2, 2B	780	3.9	287	6.1	85	92	7.0	11.8	109	177	4.2	1.96	32.0	0.71	30	26	29	5.9	1.37	0.94	5.4	2.2	0.30	10.8	—	741	717
2, 3	710	4.6	292	6.0	87	102	9.6	14.4	128	202	4.2	1.47	19.0	0.66	26	24	23	5.1	1.20	0.80	4.7	1.8	0.28	8.0	—	724	729
2, 4	820	3.1	351	5.2	41	140	5.5	12.4	69	135	4.0	1.28	7.7	0.67	28	22	26	5.6	1.29	0.90	5.0	2.2	0.28	4.5	—	838	877
2, 5	720	3.1	316	5.0	68	145	7.6	10.2	97	158	6.3	1.30	14.0	0.77	27	24	24	5.2	1.22	0.80	4.7	2.0	0.30	6.5	—	791	789
4	790	3.2	207	7.7	153	—	19.7	38.0	392	343	24.1	58.00	64.0	3.40	34	33	—	6.2	1.54	0.96	—	2.0	—	46.0	2.8	572	517
5	780	3.2	193	8.2	199	—	20.8	33.0	281	288	20.8	28.90	58.0	2.12	30	28	—	5.6	1.37	0.84	—	1.8	—	54.0	3.3	545	482
6	840	3.6	291	6.3	96	—	6.6	12.2	118	187	4.5	2.00	37.0	0.87	28	27	—	5.5	1.28	0.90	—	1.9	—	11.9	2.9	722	727
7	910	2.8	243	8.0	170	—	13.1	26.9	239	301	11.8	22.10	68.0	1.74	38	38	—	7.5	1.77	1.11	—	2.3	—	44.0	2.8	600	607
8	810	3.4	307	6.8	113	—	9.2	16.8	162	218	7.6	3.20	46.0	1.19	27	30	—	5.3	1.25	0.83	—	1.7	—	17.8	2.7	680	767
9	890	2.9	252	7.6	161	—	19.4	38.0	309	370	20.5	31.00	63.0	3.00	31	29	—	5.8	1.32	0.88	—	1.9	—	52.0	2.7	570	629
10	880	2.7	242	8.0	145	—	8.6	19.6	235	235	6.3	18.80	70.0	1.08	36	35	—	7.0	1.65	1.12	—	2.3	—	27.2	2.6	625	604
11	720	3.4	280	6.8	126	—	6.7	14.7	160	207	4.5	3.20	55.0	0.82	25	27	—	5.1	1.18	0.78	—	1.6	—	15.8	2.6	699	699
465A-1	13,400	3.5	264	7.4	146	—	29.7	77.0	461	318	33.8	2.10	57.0	7.30	15	5	—	2.5	0.66	0.42	—	1.4	—	15.6	2.9	685	659
465A-2	11,100	4.7	356	4.0	49	—	6.2	6.3	84	82	6.3	0.55	54.0	5.00	19	5	—	3.2	0.83	0.57	—	1.9	—	12.9	1.3	856	889
465A-3	5,600	—	416	0.9	4	—	0.2	1.5	14	9	0.8	—	34.0	—	12	2	—	1.8	0.47	0.35	—	1.4	—	<0.7	0.1	979	1,040

*Concentrations in $\mu\text{g per g}$ except Fe, K, Ca and CaCO_3 , mg per g, and Ir, ng per g.

†Midpoint heights are relative to the value assigned the lowest sample—see text.

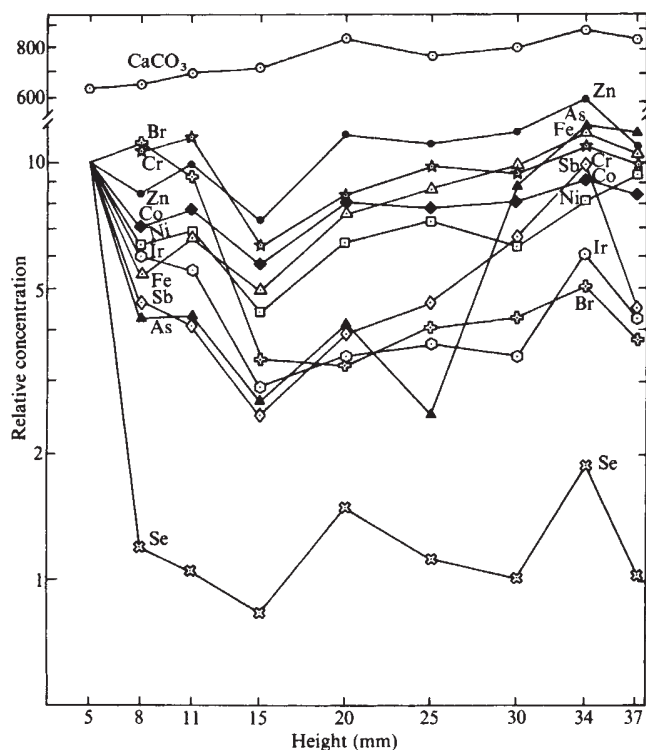


Fig. 1 Abundance profiles for 10 elements in sample 1 from the Stevns Klint fish clay. Data are on a carbonate-free basis and normalized to give sample 1, 1 a value of 10. Only Se is severely depleted up the section and only Ir, As, Sb, and Br also are depleted significantly in most samples. In general, there is little difference between the base and top of the section.

Siderophile concentrations are high in all levels of the fish clay relative to those in normal pelagic clays. The fine lamination of the clay indicates an absence of bioturbation and that there must have been a rather continuous source of siderophile-rich material. The depositional environment here into shallow basins between mounds of Maastrichtian chalk⁵, might have favoured concentration of relatively dense mineral phases. Thus, if Ir is concentrated in such phases these depressions might have become traps for siderophile-rich materials whereas clay minerals were preferentially swept towards lower regions on the continental shelf.

An extraterrestrial source of the siderophiles?

It has been suggested^{1,3,4} that the worldwide enhancement in siderophiles in the CT boundary sediments resulted from the accretion of extraterrestrial material. Alvarez *et al.*¹ inferred an extraterrestrial origin of the Ir anomaly mainly because it is worldwide, coincides with the time of the Cretaceous extinctions, and because asteroidal impacts of the required magnitude should occur about every 100 Myr. They argued against reduction of Ir from the oceanic reservoir, or a worldwide reduction in pelagic sedimentation rates. Smit and Hertogen⁴ accepted an extraterrestrial origin primarily because terrestrial sources appeared inadequate to account for the high Ir and Os at the boundary. Ganapathy³ opted for an extraterrestrial source because common terrestrial rocks have very low siderophile concentrations, and relative abundance patterns in terrestrial rocks and ores are grossly different from the roughly chondritic pattern found in the Stevns Klint clay.

An important observation is that pyrite, FeS₂, is a ubiquitous component of the CT boundary layers⁸. The formation of pyrite in marine sediments requires reducing conditions and high H₂S concentrations⁹. In basins like the Norwegian-Danian Trough where the Stevns Klint fish clay formed, these conditions generally reflect the presence of stagnant water in closed basins.

How the required conditions are generated in the deep ocean, for example, at DSDP site 465, is not clear. As a trial hypothesis, we suggest that the reducing conditions are directly associated with the extinction of Cretaceous phytoplankton, and perhaps even resulted from the excess of dead organisms; reducing conditions may have been present in the oceans simultaneously throughout the world.

In fact, such reducing conditions may be necessary to avoid severe fractionation of extraterrestrial siderophiles before deposition. The oceanic residence time of an element is defined as the amount present divided by the steady-state injection or deposition rate. In present-day steady-state conditions, siderophile residence times vary by factors of at least 10³. For example, published¹⁰ residence times are 0.20 kyr for Fe, 30 kyr for Co, 90 kyr for Ni and 120 kyr for Au. If these residence times are largely established by kinetic processes¹¹ such as ingestion by organisms, injection of large amounts of material having a solar Au/Fe ratio into the ocean would result in the first material to fall out having a Au/Fe ratio 600 times smaller than the solar ratio, whereas that in the sediments depositing a few kyr later would be much higher than the solar value.

Long residence times would have to be shortened to values comparable with the depositional periods of the CT sediments and reducing, high sulphide conditions offer plausible circumstances for this to occur. But if the residence times of siderophiles are reduced by a large factor, the deposition rate of oceanic siderophiles of terrestrial origin will be increased by a similar factor. Therefore, any deposition of relatively unfractionated extraterrestrial siderophiles is likely to be accompanied by an enhanced deposition of terrestrial siderophiles, the enhancement factor being roughly proportional to the steady-state oceanic residence time. This exacerbates the problem of separating the terrestrial and extraterrestrial signals.

Figure 2 shows our siderophile data for the six Stevns Klint samples studied by RNAA, and the INAA data available for the DSDP core 465A. In the upper portion element/Ni ratios are normalized to those in CI carbonaceous chondrites, in the lower portion to mean abundances in group IIAB iron meteorites. The following, very approximate, background concentrations based on intercepts on element-Ir diagrams were subtracted from the Stevns Klint samples: 10 mg per g Fe, 20 µg per g Co and 110 µg per g Ni. We had no basis for estimating background levels for the DSDP core. Elements are arranged in order of decreasing nebular condensation temperature^{12,13}.

Table 2 Concentrations* determined by RNAA of seven elements in the Stevns Klint fish clay

	Ir	Os	Re	Pt	Pd	Au	Ge
4	42	32	9.7	60	44	5.9	—
9	45	60	11.7	45	41	5.6	1.00
10	29	30	5.7	43	30	3.4	0.98
Allende	730	828	71.0	1,820	705	145.0	22.00
5	46	60	11.3	69	37	6.6	0.90
7	34	53	4.3	39	39	5.2	0.77
8	18	28	2.4	23	21	3.7	0.81
Allende	730	828	70.0	1,530	705	145.0	19.00

*Concentrations in ng per g except Ge, µg per g. Each neutron activation run consisted of three clay samples and the Allende CV chondrite.

The Stevns Klint samples show significant fractionations relative to any known meteorites. Nickel-normalized abundances of Os and Ir are about 5 times higher than CI values; they are ~3 times higher than the highest chondritic abundances, those in CV chondrites. Although our Re/Ni ratios are about a third of those of Ganapathy³, our Re abundance ratios are still about double those for Os and Ir.

If Ni, our normalizing element, were 2.5–3 times higher, the pattern would be generally similar to that of CV chondrites. Although our inferred background Ni value of 110 µg per g may be too high, subtracting no background at all would only increase the average Ni concentration by a factor of ~1.6, much

less than required. Moreover, there are no unique background values that would generate constant ratios of other siderophiles to Ir. If a CI pattern were initially present, then some terrestrial fractionation has surely occurred.

The estimated mean composition of group IIAB iron meteorites shows an enrichment of refractory siderophile (Os, Ir) abundances by about 3 times relative to those in CI chondrites. As shown in Fig. 2b, normalizing to IIAB leads to mean abundance ratios closer to unity than does the CI normalization in Fig. 2a. This conflicts with Ganapathy's³ inference that the impacting body was not an iron meteorite because 'most iron meteorites (with the exception of group I) show highly variable and noncosmic elemental ratios'. Ganapathy is right that individual, *m*-sized irons often show extreme fractionations associated with fractional crystallization. However, mean siderophile abundance patterns inferred by Willis¹⁴ for entire cores are generally similar to those characteristic of chondrites.

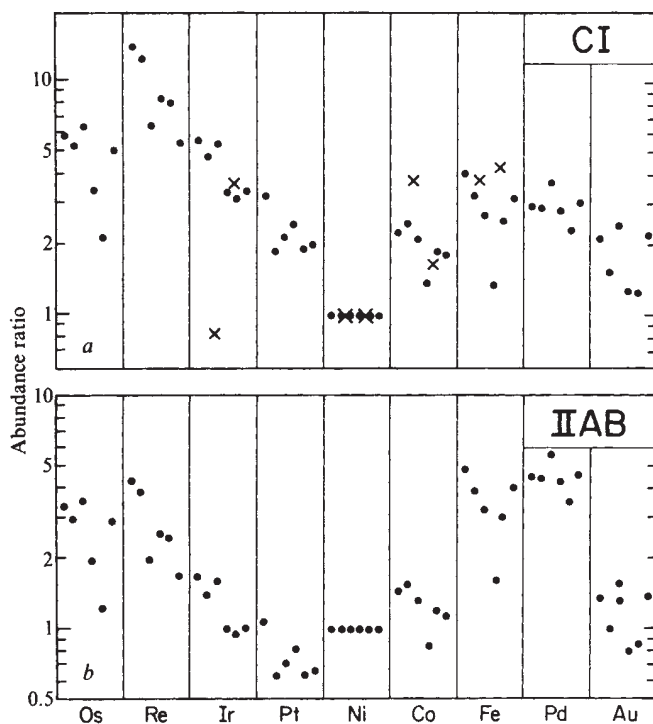


Fig. 2 Ni-normalized abundance ratios for nine siderophile elements plotted relative to a, CI and b IIAB. The six fish clay samples (●) from left to right are 5, 9, 7, 10, 4 and 8. Plot a shows the INAA data for Pacific core 465A-1 (left side) and 465A-2 (right side). Note that the fractionated patterns can be fit with both meteorite classes. The siderophiles Ge and Sb were not plotted because of the large amount of scatter introduced by correction for terrestrial materials.

A different fractionation pattern is observed in the Central Pacific sample 465A-1 (Fig. 2a). The abundance ratio of Ir is now 0.8, about 5 times lower than in the Stevns Klint clay; the Fe and especially the Co abundance ratios are even higher than in the Danish samples. Sample 465A-2 contained much less pyrite and siderophile abundances similar to the Stevns Klint sample. Perhaps the key message is that relative fractionation of siderophiles has occurred, and that the resulting patterns differ from location to location.

We can now consider whether the CT boundary siderophiles are of extraterrestrial origin. Significant fractions of some siderophiles are surely terrestrial, especially those that have relatively high oceanic residence times. The elements showing the highest abundance ratios in the Stevns Klint fish clay are Os and Ir. Alvarez *et al.*¹ report a concentration of Ir of $< 4 \times 10^{-13}$ g per g in seawater that together with a mean ocean

depth of 3,700 m and a typical pelagic deposition rates of $\sim 10^{-13}$ g Ir $\text{cm}^{-2} \text{yr}^{-1}$ (ref. 7) suggest a residence time of $\leq 10^6$ yr. It is important to establish this residence time or at least to lower the upper limit. If as seems likely the Ir residence time is comparable with or lower than the ~ 100 kyr values of Ni and Au, the enhanced Ir/Au and Ir/Ni ratios in the CT boundary sediments require that most of the Ir be of extraterrestrial origin.

An asteroidal impact?

Alvarez *et al.*¹, Smit and Hertogen⁴ and Ganapathy³ interpret the high CT boundary siderophile contents as being caused by the impact of a large (~ 10 km) asteroid. Hsü⁵ favours a cometary impact, based on the assumption that cometary ices will include chemicals toxic to phytoplankton. Comets apparently consist of roughly equal parts of ice and stony material¹⁶, thus the mass of impacting cometary material required to supply the observed accumulation of Ir would be twice that for a chondritic asteroid.

These hypotheses have serious problems in explaining the high Ir concentrations in the sediments. Table 3 lists maximum Ir concentrations observed in the four CT boundary sediments for which data are available. Because a negligible portion of the Ir is likely in the CaCO_3 , the relevant data are those corrected for CaCO_3 . The final column gives the extraterrestrial component (ETC), the ratio of Ir in the carbonate-free sediment to that in water-free CI chondrites¹⁷.

Alvarez *et al.*¹ quote the mass of impact ejecta as ~ 60 times the mass of the projectile (R.A.F. Grieve unpublished). In fact, the ratio depends on the impact velocity (Grieve's estimate seems reasonable for ~ 20 km s^{-1}), and the size fraction of the ejecta and projectile material. In the fine fraction of ejecta taken near the rims of fresh lunar craters the ratio is far higher; 4.2 ng per g Ir (ref. 18) in the $< 4\text{-}\mu\text{m}$ fraction of immature soil 61220 (Table 3), implies a ratio of lunar to chondritic material of > 100 . The two other lunar soils¹⁹ listed in Table 3 are mature, that is, their siderophile contents have been enhanced by the continued accretion of projectiles that did not penetrate the regolithic debris layers on the lunar surface. However, we doubt that, following the hypothetical CT event, the ratio in the $\leq 1\text{-}\mu\text{m}$ size fraction having an appreciable stratospheric residence was much less than that in the fine fraction of soil 61220, although we cannot rule out this extreme fractionation.

Examination of the extraterrestrial components (ETC) listed in Table 3 shows that the Danish, Spanish and Central Pacific values are much higher than expected if the projectile had a chondritic composition and the ejecta/projectile ratio were 60. Further, the terrestrial component of the CT boundary sediments is unlikely to consist entirely of ejecta; if half consisted of non-ejecta detrital material, the anticipated ETC would be only $\sim 0.8\%$.

Our rare-earth data (Table 1) can be used to examine whether the CT boundary sediment had the same composition worldwide, as would be expected were it composed entirely of ejecta. In fact both the concentrations and the patterns are significantly different: relative to Stevns Klint, DSDP core 465A has half the La concentration, La/Ce ratios are ~ 3 times higher, and La/Yb ~ 1.5 times lower. The Danish sample resembles the average ocean sediment²⁰ as well as various continental volcanic rocks; the Ce depletion in the Pacific sediment is similar to but less pronounced than the depletion in ocean water. Alteration by interaction with the ocean water during sedimentation is unlikely, because the amount of La in 3 mm of our Pacific sediment exceeds by a factor of 13 that in the overlying 2.5 km of water. We conclude that a substantial non-ejecta contribution is present in one or both sediments.

We have shown that the siderophile pattern is consistent with the projectile being an iron meteorite. The siderophile content of a metal-sulphide core of an asteroid formed from relatively reduced chondritic materials such as H-group chondrites is about 3–4 times greater, that from more oxidized chondritic materials such as L chondrites ~ 6 –7 times greater than those in

a chondritic projectile (a large fraction of all siderophiles except Fe enter the cores in both cases). If we neglect the high ETC of the Stevns Klint sample on the assumption that it reflects selective enrichment of ore minerals, and assume that the non-ejecta component of the other boundary sediments is ≤ 0.5 , a metallic core of an oxidized parent body might just account for Spanish and Pacific ETC values. About 20% of the fireballs observed by the Prairie Network enter the Earth's atmosphere at velocities $\leq 15 \text{ km s}^{-1}$ (ref. 21); if the hypothetical CT object had such a low impact velocity the ejecta/projectile ratio might be ~ 30 , half Grieves's value, and more in keeping with the high ETC values. Impact by interstellar comets or asteroids²² would tend to result in impact velocities $> 20 \text{ km s}^{-1}$.

It is very unlikely that the siderophile enrichment in the CT boundary sediments resulted from the impact with the Earth's surface of a comet or of an asteroid having a chondritic composition. The data seem consistent with the impact of a metallic core, but only if the asteroid were relatively oxidized, the impact velocity were $\leq 20 \text{ km s}^{-1}$ and if the non-ejecta fraction in the carbonate-free sediment were $\leq 50\%$.

Extraterrestrial materials not impacting the Earth's surface

If the material accreted at the end of the Cretaceous were finely divided, it would be slowed to low velocities in the Earth's atmosphere. Because no ejecta would be produced, the high ETC values would not present a problem.

The most plausible source of finely divided material is an interstellar cloud. It is estimated that the Solar System encounters such clouds each time it passes through an arm of the Galaxy, roughly once each 100 Myr (refs 22–25). The amount of Ir that could be accreted depends on the cloud's density and size.

Table 3 Maximum Ir concentrations* observed in samples from the CT boundary at four worldwide locations and the finest fractions of three lunar soils^{14,15}

Location	Whole rock	Whole rock – CaCO ₃	Extraterrestrial components† (%)
Denmark	54	120	21
Italy	5.5	9.1	1.6
Spain	27	50	9
Central pacific	13	40	7
15100 (<7 μm)	19	19	3
61200 (<4 μm)	4.2	4.2	0.7
66080 (<7 μm)	22	22	4

*In ng per g.

†Extraterrestrial components (ETC) were obtained by dividing these concentrations by those in water-free CI chondrites (Ir = 580 ng per g).

A second-order effect is the dependence of the accretion rate on its vectorial velocity, the enhancement factor being approximately $[1 + (V_{\text{esc}}^2/V_{\text{rel}}^2)]$ where V_{rel} is the relative velocity and V_{esc} is 11.2 km s^{-1} , the escape velocity from the Earth. The relative velocity is commonly estimated to be $\sim 20 \text{ km s}^{-1}$. If the velocity is in the range $20\text{--}40 \text{ km s}^{-1}$ and the vector is near the orbital plane of the earth, a significant enhancement could occur during some of the Earth's orbit, but the mean enhancement integrated around the orbit is not likely to be as large as a factor of 2. Gravitational focusing by the Sun would significantly enhance the solar accretion rate, but the enhancement of the terrestrial accretion rate would be much less. We will assume no enhancement.

Relative to Si = 10^6 atoms, the solar abundance of H is 2.2×10^{10} (ref. 26); our data on CI chondrites indicate an Ir abundance of 0.64. We will assume that these values are appropriate to interstellar clouds in our portion of the Galaxy. If the relative amount of terrestrial Ir was small at all locations, published data^{1,3,4} indicate a mean worldwide accumulation

within the range $8\text{--}80 \times 10^{-9} \text{ g cm}^{-2}$. An accumulation of $2 \times 10^{-8} \text{ g cm}^{-2}$ Ir could result from the passage of the Earth through a cloud having a mean density of $10^5 \text{ H atoms cm}^{-3}$ and a diameter of $8.6 \times 10^{19} \text{ cm}$ ($= 5.7 \times 10^6 \text{ AU} = 28 \text{ pc}$), an order of magnitude greater than that estimated for the Orion 1 cloud²⁷, one of the larger and denser interstellar clouds in our neighbourhood. Thus the probability of passage through such a dense region is probably considerably less than once each 100 Myr. Further, to account for the narrow Ir depth profiles in sediments such as that at Caravaca⁴, the bulk of the interstellar material needs to have been accreted from regions of exceptionally high density ($\geq 10^6 \text{ H atoms cm}^{-3}$), or sedimentation rates must have been $\leq 10^{-7} \text{ m yr}^{-1}$, the lowest values observed in abyssal clays today.

Additional arguments against accretion of extrasolar material are a low ^{244}Pu content ($^{244}\text{Pu}/\text{Ir}$ ratio $\leq 10^{-4}$, see ref. 1) and the absence of measurable variations in $^{191}\text{Ir}/^{193}\text{Ir}$ and $^{184}\text{Os}/^{190}\text{Os}$ ratio^{1,3,4}. However, studies of the Allende chondrite showed no variations in $^{184}\text{Os}/^{190}\text{Os}$ either in the refractory inclusions having anomalous O isotope compositions or in chromite separates having isotopically anomalous Xe (ref. 28). The absence 4.5 Gyr ago of such variations in materials of diverse isotopic composition suggests that isotopic variations of Os and perhaps Ir as well may also have shown little variation in our part of the Galaxy 65 Myr ago. If the $^{238}\text{U}/\text{Ir}$ atom ratio in the interstellar cloud was equal to the solar value of about 0.03 present at the time the Solar System formed, the $^{244}\text{Pu}/^{238}\text{U}$ ratio in the cloud was $\leq 3 \times 10^{-3}$, about 5 times lower than that initially present in the oldest meteorites²⁹. The mean $^{244}\text{Pu}/^{238}\text{U}$ surely varies by factors greater than 5 from cloud to cloud and possibly even within clouds, thus this discrepancy is not severe enough to require rejection of the interstellar cloud accretion model.

As an alternative to interstellar matter, material could enter the Earth's atmosphere as relatively large bodies but disintegrate because of low inherent strength, as thought to have occurred during the Tunguska event. De Laubenfels³⁰ first suggested that a Tunguska-like event could have triggered the Cretaceous extinctions. The Tunguska projectile was relatively small³¹ depositing an energy of $4 \pm 2 \times 10^{23} \text{ erg}$, corresponding to $2 \times 10^{11} \text{ g}$ if the velocity of 20 km s^{-1} is assumed. This is $\sim 10^7$ times smaller than the amount of material inferred from Ir in the European or Pacific CT boundary sediments. The key question is whether in the $\sim 5 \text{ s}$ atmospheric transit time a $10^{17}\text{--}10^{18} \text{ g}$ object would break up completely. We doubt that it would and suggest that an object in a non-intersecting orbit broke up into a large number of smaller objects as it passed inside the Roche limit, with a portion of the debris entering new orbits that intersected the Earth's surface.

The Tunguska object has been interpreted as a cometary remnant. In fact, Kresak³² associated it with Comet Encke. Many comets are structurally weak, as indicated by the fact that at least 10 have broken up without any apparent cause^{33,34}. It seems plausible that such weak comets would undergo breakup inside the Roche limit.

About half the meteoritic material entering the Earth's atmosphere seems to be cometary²¹, suggesting that a super Tunguska event of the sort we describe is about as likely as the impact of an asteroid composed of tough materials, perhaps amounting to one event each 100 Myr. In contrast, passage through an interstellar cloud having the required column density would seem to be a far rarer event. We favour the super-Tunguska mechanism.

Extinction mechanisms

Although infall of a comet may have marked the end of the Cretaceous, we are skeptical about Hsü's¹⁵ marine extinction mechanism, poisoning by toxic cometary substances. Although HCN, CH_3CN , and CO are certainly toxic for many organisms, the bulk of these would have been oxidized in the fireball that accompanied the cometary accretion. Further, such poisons

would not be selective among different planktonic species and would certainly have also had a deleterious effect on the benthic realm.

We suggest that cometary infall-related marine extinctions resulted either from thermal stress^{35,36} or from the opacity of the stratosphere that would accompany the infall of 10^6 – 10^7 times the mass of Tunguska¹. Total darkness for a few days or a week would lead to the death of most phytoplankton, and a domino effect on the many organisms higher in the food chain.

We are also skeptical about the suggestion¹ that stratospheric opacity lasting for several years was responsible for the death of the dinosaurs and other large land animals. The estimate¹ of 5×10^{18} g of ejecta injected into the stratosphere corresponds to the total mass of the atmosphere above ~48 km, 10% of the atmosphere above ~31 km, and 1% above ~16 km. To keep such large amounts of material suspended, it would have to behave as elastically as gas molecules. In fact, it would probably stick on impact, the coagulation following second-order kinetics. For 2–3 years following the eruption of Krakatoa in 1883 its stratospheric dust produced a reddish-brown solar corona that could be seen in favourable conditions³⁷. If favourable conditions were required to see the Krakatoa dust and if second-order kinetics controlled its fallout, clearly no matter how much dust you inject into the stratosphere, within a few months or less the level will drop to a value no more than a few times greater than that present immediately following the Krakatoa explosion. Stratospheric opacity could appreciably reduce photosynthesis by land plants for no more than one season, and probably a much shorter period.

It seems more likely that a major input of energy, whether by asteroidal impact or a super Tunguska, would lead to extensive destruction of forests by the blast wave and accompanying forest fires, as occurred within about a 100 km² area about the Tunguska epicentre³⁸. Clearly no land animal larger than 25 kg (ref. 39) could have continued to survive in the scorched area near Tunguska.

Another extinction mechanism is required if the siderophiles were accreted from an interstellar cloud. In this case accretion to the solar surface would result in the conversion of gravitational energy to heat, with an increase in the luminosity, and extinctions due to thermal stress^{35,40}. A cloud density of 10^5 H cm⁻³ and a relative velocity of 20 km s⁻¹, as used in our estimate, would increase the solar luminosity by 1% (ref. 41). Further, the increase depends on V_{rel}^{-3} , thus a relative velocity about 25% lower would double the effect. One can easily imagine that as a result of the scatter in velocities and densities within a large interstellar cloud, there is a high probability that during short periods (say 100 yr) the solar luminosity would be increased by 20% or more, resulting in a mean increase in temperature of ≥ 15 K. Additional climatic effects are associated with solar

flares and distortions of the solar and terrestrial magnetic fields while the Solar System is immersed in an interstellar cloud^{24,25}.

One interesting effect is the reduction of atmospheric O₂ as a result of the accretion of interstellar H₂ to the Earth (see also ref. 42). The estimated accumulation of Ir, 2×10^{-8} cm⁻² corresponds to 3.5 g cm⁻² of H, 14% of the amount required to fully convert atmospheric O₂ to H₂O. In contrast to CO₂, there is weak buffering on the atmospheric O₂ content; the level at any time represents a dynamic balance between photosynthetic production and various sinks; adjustment to shifts in the equilibrium abundance requires a few Myr (ref. 43). Baur and Friedl⁴⁴ recently suggested that early Mesozoic O₂ levels were only ~60% of present day levels. If levels in the Cretaceous 100 Myr later were also low, the impact of accretion of interstellar H could be still greater than the above estimates.

Undoubtedly many other effects deleterious to one or the other families of organisms would result from immersion of the Solar System in an interstellar cloud.

Conclusions

A pyrite-bearing CT boundary clay from DSDP hole 465A in the Central Pacific has high concentrations of Ir and other siderophiles, confirming that the high siderophile contents found in shallower sediments are also found in those deposited at depths >2 km. Concentrations of 25 elements as a function of height in the finely layered CT boundary clay from Stevns Klint, Denmark show that although the highest concentrations are found in basal samples, carbonate-free concentrations of siderophiles are relatively high at all heights. The ubiquitous presence of pyrite in CT boundary sediments indicates extensive reducing conditions—these probably led to a reduction in the oceanic residence times and thus enhanced deposition for some terrestrial siderophiles. Nevertheless, we tentatively conclude that most of the Ir in the CT sediments is of extraterrestrial origin. Concentrations of Ir are higher than expected from the surficial impact of a comet or an asteroid of chondritic composition; a metal-sulphide asteroidal core is marginally possible, but it seems more likely that the siderophile-rich material reached terminal velocity in the atmosphere. There is a small possibility that this was dust from an interstellar cloud, but it is more probable that the near-Earth breakup of weak cometary materials produced a series of Tunguska-like events all around the globe.

We thank H. Tappan and A. R. Loeblich for discussions and for the Stevns Klint samples. The collection of the DSDP samples was supported by the NSF and L. Garifal and W. R. Riedel provided these samples. We thank F. Asaro, L. W. Alvarez, P. R. Weissman and J. Hertogen for helpful comments and J. Pai and A. Young for technical assistance.

Received 8 October; accepted 11 November 1980.

- Alvarez, L. W., Alvarez, W. Asaro, F. & Michel, H. V. *Science* **208**, 1095–1108 (1980).
- Alvarez, W., Alvarez, L. W., Asaro, F. & Michel, H. V. *Geol. Soc. Am. Abstr. Prog.* **11**, 378 (1979).
- Ganapathy, R. *Science* **209**, 921–923 (1980).
- Smit, J. & Hertogen, J. *Nature* **285**, 198–200 (1980).
- Christensen, L., Fregerslev, S., Simonsen, A. & Thiede, J. *Bull. geol. Soc. Denmark* **22**, 193–212 (1973).
- Nadkarni, R. A. & Morrison, G. H. *Analyt. Chem.* **46**, 232–236 (1974).
- Crockett, J. G. & Kuo, H. Y. *Geochim. cosmochim. Acta* **43**, 831–842 (1979).
- Percival, S. F. & Fischer, A. G. *Evolut. Theory* **2**, 1–35 (1977).
- Berner, R. A. *Principles of Chemical Sedimentology* (McGraw-Hill, New York 1971).
- Brewer, P. G. in *Chemical Oceanography* (Riley, J. P. & Skirrow, G.) 415–495 (Wiley, New York, 1975).
- Broecker, W. S. *Quat. Res.* **1**, 188–207 (1971).
- Wai, C. M. & Wasson, J. T. *Nature* **282**, 790–793 (1979).
- Palme, H. & Wlotzka, F. *Earth planet. Sci. Lett.* **33**, 45–60 (1976).
- Willis, J. thesis, Univ. California, Los Angeles (1980).
- Hsü, K. J. *Nature* **285**, 201–203 (1980).
- Delsemme, A. H. *Comets, Asteroids, Meteorites* (ed. A. H. Delsemme) 3–12 (University Toledo, 1977).
- Wasson, J. T., Boynton, W. V., Chou, C.-L. & Baedeker, P. A. *The Moon* **13**, 121–141 (1975).
- Boynton, W. V., Chou, C.-L., Bild, R. W., Baedeker, P. A. & Wasson, J. T. *Earth planet. Sci. Lett.* **29**, 21–33 (1976).
- Boynton, W. V. & Wasson, J. T. *Geochim. cosmochim. Acta* **41**, 1073–1082 (1977).

- Haskin, L. A. & Paster, T. P. *Handbook on the Physics and Chemistry of Rare Earths* (eds Gschneidner, K. A. Jr & Eyring, L.) 1–80 (North Holland, Amsterdam 1979).
- Wetherill, G. W. A. *Rev. Earth Planet. Sci.* **2**, 303–331 (1974).
- Napier, W. M. & Clube, S. V. M. *Nature* **282**, 455–459 (1979).
- McCrea, W. H. *Nature* **225**, 607–609 (1975).
- Talbot, R. J., Butler, D. M. & Newman, M. J. *Nature* **262**, 561–563 (1976).
- Begelman, M. C. & Rees, M. J. *Nature* **261**, 298–299 (1976).
- Ross, J. E. & Aller, L. H. *Science* **191**, 1223–1229 (1976).
- Werner, M., Becklin, E. & Neugebauer, G. *Science* **197**, 723–732 (1977).
- Takahashi, H., Higuchi, H., Gros, J., Morgan, J. W. & Anders, D. *Proc. natn. Acad. Sci. U.S.A.* **73**, 4253–4256 (1976).
- Podosek, F. A. A. *Rev. Astr. Astrophys.* **16**, 293–334 (1978).
- De Laubenfels, M. W. J. *Paleontol.* **30**, 207–218 (1956).
- Hunt, J. N., Palmer, R., Penney, W. *Phil. Trans. R. Soc. A252*, 275–315 (1960).
- Kresak, L. *Bull. astr. Instr. Czech.* **29**, 129–134 (1978).
- Stefanik, R. P. *Mem. Soc. R. Sci. Liège, Ser. 5* **12**, 29–32 (1966).
- Sekanina, Z. *Icarus* **30**, 574–594 (1977).
- Opik, E. J. *Ir. astr. J.* **5**, 34 (1958).
- Emiliani, C. *EOS* **61**, 505–506 (1980).
- Humphreys, W. J. *Physics of the Air* (McGraw-Hill, New York, 1940).
- Krinov, E. L. *Giant Meteorites* (Pergamon, Oxford, 1966).
- Russell, D. A. A. *Rev. Earth planet. Sci.* **7**, 163–182 (1979).
- McLean, D. M. *Science* **201**, 401–406 (1978).
- Hoyle, F. & Lyttleton, R. A. *Proc. Camb. Phil. Soc. Math. Phys. Sci.* **35**, 405–415 (1939).
- Alvarez, L. W., Alvarez, W., Asaro, R. & Michel, H. V. *Univ. Calif. Rep. LBL-9666* (Lawrence Berkeley Laboratory, 1979).
- Walker, J. C. G. *Am. J. Sci.* **274**, 193–214 (1974).
- Baur, M. E. & Friedl, R. R. *Am. Ass. Adv. Sci. Selected Symp.* **28**, 254–286 (1980).

Basement heat flow and metalliferous mineralization in England and Wales

G. C. Brown*, J. Cassidy*, E. R. Oxburgh†, J. Plant‡, P. A. Sabine§ & J. V. Watson||

* Department of Earth Sciences, The Open University Milton Keynes MK7 6AA, UK

† Department of Mineralogy and Petrology, Downing Place, Cambridge CB2 3EQ, UK

‡ Institute of Geological Sciences (NERC), 154 Clerkenwell Road, London EC1R 5DU, UK

§ Institute of Geological Sciences (NERC), Exhibition Road, London SW7 2DE, UK

|| Department of Geology, Imperial College of Science, South Kensington, London SW7 2BP, UK

Many examples of epigenetic metalliferous mineralization in central and northern England occur within a province now recognized to be characterized by high heat flow. Our analysis of this correlation leads us to conclude that Caledonian intrusive bodies, with high heat production and thermal conductivity, focused the development of hydrothermal systems responsible both for some types of post-Carboniferous mineralization and for present-day high heat flow.

TWO recent developments in the exploration for geothermal energy and metalliferous mineralization in Britain have highlighted a possible connection between the present-day heat production of rocks (due mainly to the long-lived isotopes ^{238}U , ^{235}U , ^{232}Th and ^{40}K) and the distribution of epigenetic metalliferous mineralization dating from Hercynian and post-Hercynian times.

First, two high heat flow provinces characterized by surface heat flow $>60\text{ mW m}^{-2}$ have been identified in Britain (Fig. 1a)^{1,2}. The northern England high heat flow province (hereafter called the Province) extends as a southeasterly trending zone at least 100-km wide from southern Scotland across the Pennines through Yorkshire to the East Midlands; the south-west England province extends eastwards from Cornwall to Hampshire. Surface heat flow in these provinces is linearly correlated with heat production in crystalline rocks and low-grade metamorphic rocks of the pre-Variscan or Variscan basement, and the provinces show a strong spatial association with the known ore fields of the Lake District, the north and south Pennines (including Derbyshire) and Cornubia (Fig. 1b). Although the mineralized zone in the South Pennines lies on the western margin of the Province, various authors (see refs 3, 4) have indicated that the source of the mineralizing fluids lies to the east. Many occurrences of epigenetic mineralization may, therefore, be related to present-day zones of high heat flow and, where data are available, to above-average basement heat production⁵.

Second, it has become evident that the British granites which are associated with vein mineralization contain unusually high levels of heat-producing elements⁶. These granites seem to have influenced the circulation of hydrothermal fluids not only at the time of emplacement but also, by virtue of their high heat production and relatively high thermal conductivity, long after the period of their consolidation. For example, in south-west England, the Hercynian granites are associated with U–Sn–Pb mineralization, followed by Pb–Zn mineralization at lower temperatures, both dating roughly from the time of emplacement⁷; in addition, U mineralization seems to have continued into Tertiary times⁸. In northern England, where no granites younger than Caledonian are known, mineralization is often located in Carboniferous or post-Carboniferous rocks which post-date the initial cooling of the intrusions, but is known to be spatially related to intrusions under the Alston and Askrigg Blocks and in the Lake District. The model proposed here for the Province whereby hydrothermal convection is maintained by high heat production in granitic rocks has recently been proved quantitatively viable for the origin of uranium deposits in New Hampshire⁹.

Here we examine possible implications of these relationships which provide the basis for current geophysical and geochemical research being sponsored by NERC as part of the coordinated UK Deep Geology programme. Consideration is given to the processes of mineralization in the Province and to the possibility that concealed mineralization may be detectable.

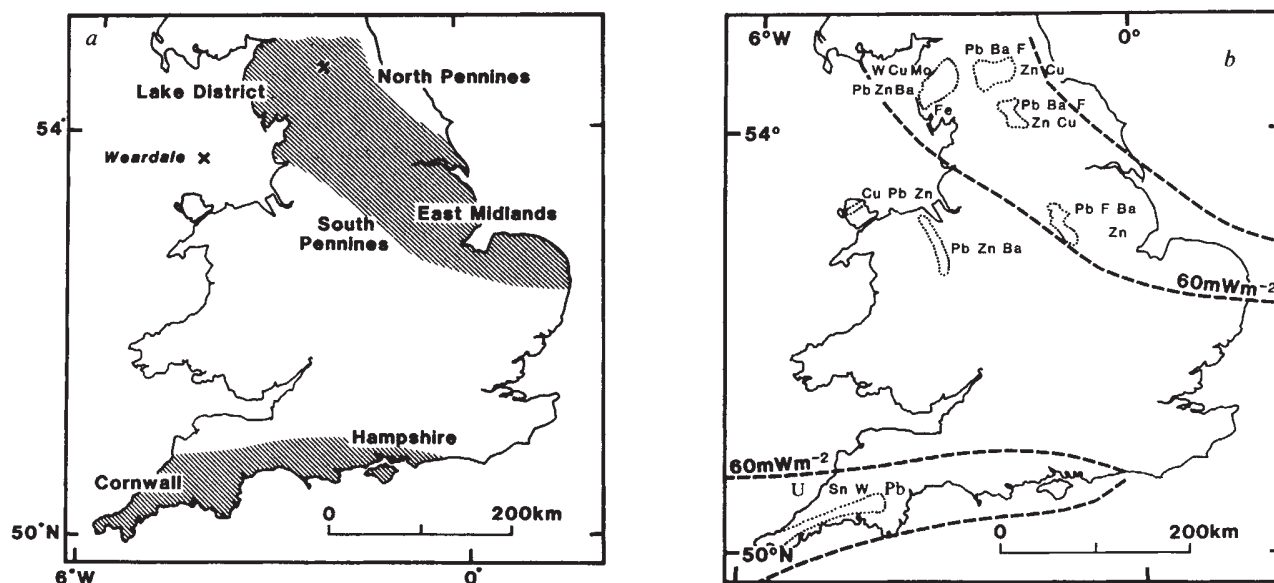


Fig. 1 a, Provinces of high heat flow in central and southern Britain; heat flow in the shaded areas exceeds 60 mW m^{-2} indicated by dashed lines in b, which also records the spatial occurrence of some important centres of hydrothermal mineralization in England and Wales (heat flow data from refs 2, 16, 22, 23, 31).

Heat sources and mineralization

Regional geology suggests that the Province cannot be directly related to the distribution of any single stratigraphical unit in the pre-Permian basement (see ref. 2). Devonian and Carboniferous rocks, though present in some areas, are often thin or absent in the northwestern part of the Province. The sub-parallel alignment, on a regional scale, of the Province and the smooth linear anomalies of deep source shown on the 1:625,000 Aeromagnetic map¹⁰ suggests a possible connection with pre-Carboniferous basement structure. Over much of its extent the Province is known to coincide with a Lower Palaeozoic basement-age province¹¹; this correlation alone does not account for high heat flow which, instead, may be related to above-average heat production in granitic rocks and greenschist facies shale belts within the basement, both of which contain relatively high levels of radioelements.^{2,5,6} The significance of these relationships is discussed below with reference to three principal areas.

(1) In the Lake District, there is a clear spatial relationship between Caledonian granites and later mineralization¹²⁻¹⁴. Average heat productivities in the Shap, Skiddaw and Eskdale granites ($4.37 \mu\text{W m}^{-3}$, $2.65 \mu\text{W m}^{-3}$, $2.0 \mu\text{W m}^{-3}$, respectively) are high compared with, for example, an average value of $1.5 \mu\text{W m}^{-3}$ for Borrowdale volcanics¹⁵ and a range of 0.6 – $2.0 \mu\text{W m}^{-3}$ for low-grade metasedimentary and metavolcanic rocks beneath England and Wales¹⁶. Recent studies^{17,18} have shown that these intrusions are characterized by large negative Bouguer anomalies, indicating that large volumes of similar igneous rocks are present at a depth below the present erosion level. In geochemical and geophysical characteristics, these and other late Caledonian intrusions are comparable with the Hercynian plutons of the south-west England high heat flow province^{1,6,19}. The relatively unfractionated heat-productivity profile of the Cornubian batholith in particular (requiring that the measured heat productivities are preserved to depths approaching 15 km (ref. 16)) might also apply in the Caledonian plutons. This suggestion is consistent with mineralogical studies²⁰ indicating that, away from areas of mineralization, uranium in both the Cornubian and late Caledonian granites is held in primary mineral phases. By virtue of their high heat production and thermal conductivity, Caledonian granites beneath the Province would not only have added heat to fuel convective systems involving formation waters in Lower Carboniferous and older sediments beneath thick sedimentary cover²¹, but they might also have continued to act as significant conductive heat-transfer agents during later periods. A mean thermal conductivity of $3.3 \text{ W m}^{-1} \text{ K}^{-1}$ for comparable Cornubian granites²² is much higher, for example, than a typical value of $1.5 \text{ W m}^{-1} \text{ K}^{-1}$ for shale horizons²³. In some circumstances where upward deformation of sediments has occurred around granites, heat may have been refracted towards those granites (for details, see ref. 5).

(2) Well documented areas^{24,25} of epigenetic mineralization above buried Caledonian granites occur in the north Pennines (Fig. 1) at Weardale and Wensleydale. For example, the Rookhope borehole, drilled to investigate the coincidence of the centre of a zonal pattern of Pb–Zn–fluorite–barite mineralization in Carboniferous rocks^{26,27} and areas of strong negative Bouguer anomaly²⁸, proved the presence of the Caledonian-aged Weardale granite²⁴. Whether mineralization in the north Pennine orefield occurred during a single short-lived episode²⁹ or in repeated episodes³⁰, it certainly took place long after these ~400-Myr-old granites had crystallized: a possible link to radioactive heat production is implied. A geothermal gradient of 31°C km^{-1} at Weardale³¹ supports comparisons of the thermal properties of this granite with those of the Cornubian batholith where gradients of 27 – 35°C km^{-1} are typical²². If the model for mineralization developed for south-west England (ref. 6 and above) is applied to late Caledonian granites, particularly those which are covered by sediments, it follows that concealed mineralization may exist elsewhere in buried sediments which are younger than the granites and with which there would be a close spatial relationship.

(3) The south Pennine–East Midlands section of the Province (Fig. 1) contains the vein and stratiform Pb–Zn–fluorite mineralization of Derbyshire which occupies a tract with NNW elongation. Ford⁴ postulated a migration of brines from the North Sea basin to account for the availability of mineralizing fluids in the area. This interpretation, however, (see ref. 32) involves the proposals that the fluids travelled some 150 km up-dip in Carboniferous strata without significant heat loss and that they crossed the Askern structural high³³. Such difficulties focus attention on the possible influence of local heat sources and, by comparison with the north Pennines, Evans and Maroof³⁴ suggested that there may be Caledonian intrusive rocks beneath the East Midlands and that these might have exerted an important control. Unfortunately, present geophysical data are equivocal in respect of the presence or absence of buried granite bodies in the south Pennine–East Midlands region. The available deep boreholes penetrate pre-Carboniferous shales with moderate heat production (our unpublished data) and do not reach crystalline basement. Although there is no clear explanation for above-average heat flow in the south-west quadrant of the Province we postulate, by analogy with the Lake District–North Pennine areas, that high basement heat flow related to granites may be responsible: the geophysical implications of this suggestion are receiving further attention. However, we note that water circulating through and rising from deep aquifers beneath horizons of low thermal conductivity provides a more attractive explanation for high heat flow in some parts of the Province, such as east Lincolnshire and Yorkshire, which are underlain by deep sedimentary basins.

Prerequisites for Pennine mineralization

Epigenetic Pb–Zn vein-type and strata-bound ores of the Mississippi Valley type are characteristically of low-temperature hydrothermal origin^{35,36} and, as such, their formation requires slightly elevated thermal gradients, trapped halogen-rich aqueous fluids in permeable rocks and preferably a structural anomaly, such as a fault zone or fractured anticline, to promote fluid circulation. Fluid circulation is also favoured by sources of magmatic or tectonic heat, but these are not essential prerequisites. In the case of Pennine mineralization, it is generally agreed that Lower Carboniferous limestones and sandstones contained the trapped connate water but that the source of base metals was either within Lower Carboniferous shales or within earlier Palaeozoic metasediments^{4,30,32,34}. The metals were leached by hydrothermal brines which, from fluid-inclusion studies³², are known to have had high salinities and temperatures (~200 °C). Some $^{34}\text{S}/^{32}\text{S}$ ratios for galena³⁷ and, in particular, the quantity and the rare-earth element geochemistry of fluorite in Pennine mineralization have both been taken to indicate, however, that the genesis of these deposits could have been related to magmatic sources³⁸.

An important point of contrast between the Pennine deposits and those of both the Mississippi Valley and the Irish Pb–Zn deposits within Lower Carboniferous host rocks^{40,41} is that whereas the latter occurrences are fault-controlled³⁹, the Pennine deposits are spatially related to domes. In view of the age discrepancy mentioned earlier, this may be taken as evidence that the role of older Caledonian granites was not only to act as channelways for mineralizing fluids³⁰, but that they may have been heat-focusing agents. Flow of fluorine-rich fluids through these granites may have remobilized base metals during hydrothermal processes and may have led to continued enrichment. Differences between the sulphide assemblages of the two Pennine areas suggest a higher temperature regime in the northern (Alston) region^{42,43}, a feature which correlates with recent data⁵ on the differences in heat flow from the two blocks. In addition to possible contributions from heat-focusing granites, the flow of mineralizing fluids was controlled by faults⁴⁴ formed during differential uplift at the end of the Carboniferous, resulting in the deposition of ores at intersections with favourable sedimentary horizons.

Implications for mineral exploration

The present model is that metalliferous minerals were concentrated magmatically during a Caledonian phase and may have been redeposited in mineral veins near the contemporaneous land surface or on or just beneath a shallow sea floor if later phases of mineralization took place. The process could have continued as long as hydrothermal systems were able to circulate against the hydrostatic gradient, possibly into Tertiary times. The search for minerals should, therefore, be concentrated at and below ancient land or sea-bottom surfaces. Several lines of evidence, including studies of metamorphic facies¹⁶ and levels of granite intrusions⁴⁵, suggest that the present land surface of the northern end of the Province is close to that present at the end of the Devonian. If hydrothermal circulation is able to reach near-surface conditions^{6,7}, it follows that most epigenetic mineralization associated with areas that have been positive blocks during post-Devonian time must crop out at, or near, the surface and may be found by established methods of geochemical and geophysical exploration^{46,47}. Sedimentary basins containing Carboniferous and later strata, such as those which occupy the central and southeastern part of the Province, may contain deposits at greater depths.

Comparisons with the known history of mineralization in the south-west England high heat flow province (particularly Cornwall, Fig. 1) suggests that the possible existence of two other types of deeply buried deposit in the northern Province is worth consideration. First, in Cornwall, where there is evidence of several post-Hercynian episodes of U remobilization analogous to those discussed above, Sn-W mineralization seems to

have been related only to a high-temperature event closely linked with magmatism. Metals which were not remobilized by low-temperature hydrothermal processes may remain in similarly close association with buried Caledonian granites. Second, stratiform metalliferous deposits in Carboniferous and Permian strata may exist beneath the cratonic cover of Jurassic or younger sediments, more particularly in the southeastern part of the northern Province.

The continuation of mineralizing processes in areas containing radioelement-enriched buried granites, long after the emplacement of those intrusions (see refs 9, 24, 48) is not always widely appreciated. An interesting example of extensive hydrothermal circulation has been recognized in the vicinity of a Tertiary granitic stock near Marysville, Montana⁴⁹, where fluids which circulated in a diffuse fracture system are thought to have set up high thermal gradients near the stock. The igneous rocks themselves are too old for magmatic heat to be responsible for the present-day hydrothermal systems. Other examples of modern hydrothermal systems associated with Palaeozoic radiogenic intrusive bodies from the Eastern Appalachians have been described by Costain and others⁵⁰. Processes of mineralization associated with such hydrothermal systems are likely to be most active in provinces of high heat flow, a consideration which focuses attention on the northern England high heat flow province.

We thank G. M. Brown, F. W. Dunning and D. A. Gray for reviews. This paper is published with the permission of the director of the Institute of Geological Sciences.

Received 2 June; accepted 31 October 1980.

1. Tammemagi, H. Y. & Wheildon, J. *Geophys. J. R. astr. Soc.* **38**, 83–94 (1974).
2. Richardson, S. W. & Oxburgh, E. R. *Nature* **282**, 565–567 (1979).
3. Bullard, E. C. & Niblett, E. R. *Mon. Not. R. astr. Soc. Geophys. Suppl.* **6**, 222–238 (1951).
4. Ford, T. D. in *Handbook of Strata-bound and Stratiform Ore Deposits* Vol. 5 (ed. Wolf, K. H.) 161–195 (Elsevier, Amsterdam, 1976).
5. England, P. C., Oxburgh, E. R. & Richardson, S. W. *Geophys. J. R. astr. Soc.* **62**, 439–455 (1980).
6. Simpson, P. R., Brown, G. C., Plant, J. & Ostle, D. *Phil. Trans. R. Soc. A* **291**, 385–412 (1979).
7. Sheppard, S. M. F. *J. geol. Soc. Lond.* **133**, 573–591 (1977).
8. Darnley, A. G., English, T. H., Sprake, O., Preece, E. R. & Avery, D. *Miner. Mag.* **34**, 159–176 (1965).
9. Fehn, U., Cathles, L. M. & Holland, H. D. *Econ. Geol.* **73**, 1556–1566 (1978).
10. *1:625,000 Aeromagnetic Maps of Great Britain* (IGS, London, 1965).
11. Wills, L. J. *Mem. geol. Soc. Lond.* **8** (1978).
12. Ineson, P. R. in *Handbook of Strata-bound and Stratiform Ore Deposits* Vol. 5 (ed. Wolf, K. H.) 198–230 (Elsevier, Amsterdam, 1976).
13. Firman, R. J. in *The Geology of the Lake District* (ed. Moseley, F.) 226–241 (Yorkshire Geological Society Occasional Publ. 3, 1979).
14. Ineson, P. R. & Mitchell, J. G. *Geol. Mag.* **111**, 521–537 (1974).
15. Cassidy, J. thesis, Liverpool Univ. (1979).
16. Richardson, S. W. & Oxburgh, E. R. *J. geol. Soc. Lond.* **135**, 323–337 (1978).
17. Bott, M. H. P. *J. geol. Soc. Lond.* **130**, 309–331 (1974).
18. Locke, C. A. & Brown, G. C. *Nature* **272**, 526–528 (1978).
19. Brown, G. C., Plant, J. & Lee, M. K. *Nature* **280**, 129–131 (1979).
20. Simpson, P. R., Plant, J. & Cope, M. J. in *Geology, Mining and Extractive Processing of Uranium* (ed. Jones, M. J.) 126–139 (Instn Mining and Metallurgy, London, 1976).
21. Russell, M. J. *Trans. Instn Min. Metall. (Sect. B)* **87**, B168–B171 (1978).
22. Wheildon, J., Francis, M. F. & Thomas Betts, A. *Seminar on Geothermal Energy* Vol. 1 175–188 (Commission of the European Communities, Brussels, 1977).
23. Oxburgh, E. R., Richardson, S. W., Bloomer, J. R., Martin, A. & Wright, S. *Seminar on Geothermal Energy* Vol. 1, 155–173 (Commission of the European Communities, Brussels, 1977).
24. Dunham, K. C., Dunham, A. C., Hodge, B. L. & Johnson, G. A. L. *Q. Jl geol. Soc. Lond.* **121**, 383–417 (1965).
25. Bott, M. H. P. *Proc. Yorks. geol. Soc.* **36**, 139–168 (1967).
26. Dunham, K. C. *Q. Jl geol. Soc. Lond.* **90**, 689–720 (1934).
27. Small, A. T. *Trans. Instn. Min. Metall. (Sect. B)* **87**, B10–B13 (1978).
28. Bott, M. H. P. & Masson-Smith, D. Q. *Jl geol. Soc. Lond.* **113**, 93–117 (1957).
29. Smith, F. W. & Phillips, R. *Fortschr. Miner.* **32**, 491–494 (1975).
30. Dunham, K. C. in *The Geology and Mineral Resources of Yorkshire* (eds. Rayner, D. H. & Hemingway, J. E.) 293–308 (Yorkshire Geological Society Occasional Publ. 2, 1974).
31. Bott, M. H. P., Johnson, G. A. L., Mansfield, J. & Wheildon, J. *Geophys. J. R. astr. Soc.* **27**, 277–288 (1972).
32. Rogers, P. J. *Trans. Instn Min. Metall. (Sect. B)* **86**, B128–B132 (1977).
33. Kent, P. E. *Proc. Yorks. geol. Soc.* **36**, 127–133 (1967).
34. Evans, A. M. & Maroof, S. I. *Min. Mag. Lond.* **134**, 401–411 (1976).
35. Dozy, J. J. *Trans. Instn. Min. Metall. (Sect. B)* **79**, B163–B170 (1970).
36. Mitchell, A. H. G. & Garson, M. S. *Miner. Sci. Engng* **8**, 129–169 (1976).
37. Robinson, B. W. & Ineson, P. R. in *25th int. geol. Congr. Australia 1976 Abstr.* 1, sect. 4A, 183 (1976).
38. Russell, M. J. & Smith, F. W. *Trans. Instn Min. Metall. (Sect. B)* **88**, B30 (1979).
39. Carpenter, A. B., Trout, M. L. & Pickett, E. E. *Econ. Geol.* **69**, 1191–1206 (1974).
40. Morrissey, C. J., Davis, G. R. & Steed, G. M. *Trans. Instn Min. Metall. (Sect. B)* **79**, B174–B185 (1971).
41. Russell, M. J. *Trans. Instn Min. Metall. (Sect. B)* **83**, B128–B133 (1975).
42. Rogers, P. J. *Trans. Instn Min. Metall. (Sect. B)* **87**, B125–B131 (1978).
43. Vaughan, D. J. & Ixer, R. A. *Trans. Instn Min. Metall. (Sect. B)* **89**, B99–B109 (1980).
44. Carter, J. S. & Moore, J. Mc. *Trans. Instn Min. Metall. (Sect. B)* **87**, B90–B93 (1978).
45. Leake, B. E. in *Geol. J. Spec. Iss.* **10**, 221–248 (1978).
46. Hawkes, H. E. & Webb, J. S. *Geochemistry in Mineral Exploration* (Harper and Row, New York, 1962).
47. Bowie, S. H. U. *Proc. R. Soc.* **339**, 299–311 (1974).
48. Rich, R. A., Peterson, U. & Holland, H. D. *Hydrothermal Uranium Deposits* Vol. 6 (Elsevier, Amsterdam, 1977).
49. *Energy Research and Development Administration, Rep. No. 1 of Hot Dry Rock Assessment Panel* (ERDA 77–74, Springfield, 1977).
50. Costain, J. K., Glover, L. III & Sinha, A. K. *EOS* **61**, 1–3 (1980).

Pacifica and New Zealand: proposed eastern elements in Gondwanaland's history

P. J. J. Kamp

Earth Sciences and Antarctic Research Unit, University of Waikato, Hamilton, New Zealand

The origin of the sediments of the Canterbury Suite to the east of the Alpine Fault in New Zealand, cannot have derived from Marie Byrd Land before or during the breaking up of Gondwanaland. It is argued that the sediments may have derived from the lost continent of Pacifica, which is supposed to have separated from Gondwanaland in the late Palaeozoic. Such a mechanism can also account for the recurrence of similar quartzo-feldspathic sediments in California and thus of the Mesozoic dolerites and basalts in southern Africa, Antarctica and Tasmania.

THE source of voluminous continent-derived sandstones¹ deposited on the Pacific margin of Gondwanaland, but oceanwards of an active arc-trench environment², is a critical problem in the late Palaeozoic-Mesozoic geology of New Zealand and one which has wider implications for Pacific geology. Eastern Gondwanaland (Australia-Antarctica) is precluded as the chief source of the quartzo-feldspathic detritus, known as the Canterbury Suite, for several reasons: (1) the sedimentology of the suite indicates derivation of the sediment from a source in the east (Palaeo Pacific Ocean); (2) the bulk petrography is inconsistent with transportation through the then active volcanic arc bordering eastern Gondwanaland; and (3) alternative explanations, especially of post-depositional juxtapositioning of the Canterbury and Wakatipu Suites by strike-slip faulting subparallel to Antarctica, are excluded by the Mesozoic position with respect to Gondwanaland of the Lord Howe Rise and Campbell Plateau. I propose that this sediment was derived mainly from Pacifica³, a continent contiguous with eastern Gondwanaland until its separation from the supercontinent during the late Palaeozoic, its fragmentation in the Triassic and inclusion within the circum-Pacific Cordillera in the early Cenozoic. Pacifica may also be the hitherto unknown source of other anomalous occurrences of quartzo-feldspathic detritus around the Pacific such as the Franciscan assemblage in California⁴. Furthermore, the origin of the extensive Jurassic basalts and dolerites of southern Africa⁵, Antarctica⁶ and Tasmania⁷ may be related to the separation of Pacifica.

The source of the Canterbury Suite has been debated for many years, and this problem is considered here in three ways. (1) The nature of the Canterbury Suite and the evidence for derivation from an eastern continent is outlined. (2) A new Mesozoic reconstruction of New Zealand relative to Gondwanaland is presented; it is important to consider New Zealand's position as close as possible to the time of Canterbury Suite deposition because the position of accumulation is a constraint on potential source areas. (3) Other problems in

Pacific geology are considered to show that the Canterbury Suite is not an isolated anomaly, and that Pacifica is a general as distinct from a merely local explanation of the source problem.

The fundamental difference between previous hypotheses of the source of the Canterbury Suite and the hypothesis advanced here is one of invoking present continents (Australia-Antarctica) as opposed to a lost continent—Pacifica. In addition to providing a source, the Pacifica model provides a mechanism, through continental fragmentation, for the periodic and rapid derivation of huge volumes of quartzo-feldspathic sediment. Further, the separation and fragmentation of Pacifica are consistent with a generally accepted plate tectonic interpretation² of the arc-trench system that formed the adjacent Wakatipu Suite.

Late Palaeozoic-Mesozoic New Zealand

The late Palaeozoic-Mesozoic geology of New Zealand is dominated by the juxtaposition of two contrasting detrital mineral suites, one the western volcanoclastic Wakatipu Suite of andesitic composition deposited next to the foreland, and the other the quartzo-feldspathic Canterbury Suite deposited in the Pacific Ocean further to the east¹ (Fig. 1). These coeval suites are considered to have formed in contrasting tectonic environments and to have been rafted together in the Mesozoic which culminated in a collision—the Rangitata Orogeny—with intense deformation, metamorphism, uplift and erosion².

The Canterbury Suite comprises two-thirds of the Rangitata Sequence⁸ and has an estimated⁹ thickness of 30,000 m which is not a true stratigraphical thickness but caused by tectonic stacking or a shifting dopocentre. Derivation from a terrain of granodiorite and minor schistose and metamorphic rock is indicated by the composition of the dominant minerals and rock fragments. In particular Triassic sandstones comprise quartz (40–50%), potash feldspar (10–15%), sodic plagioclase (20%) and minor amounts of muscovite, volcanic rock fragments and chert¹. These sandstones typically average SiO₂ contents of

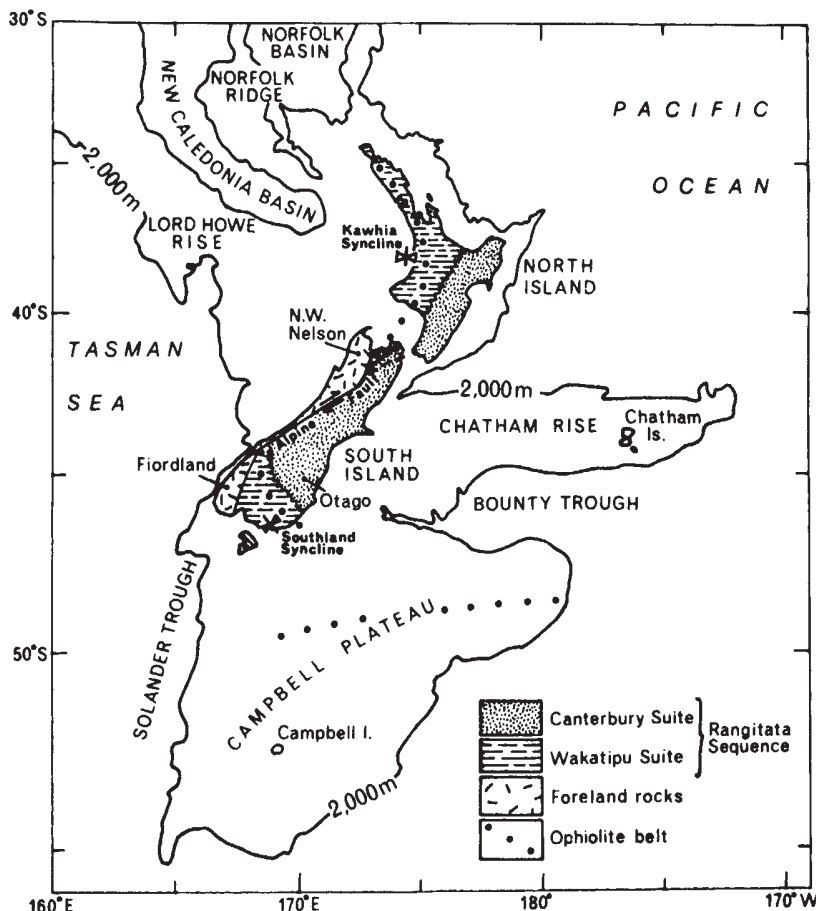


Fig. 1 A map of the New Zealand sub-continent, defined by the 2,000-m isobath, showing the on-land distribution of Palaeozoic foreland rocks¹² and the late Palaeozoic-Mesozoic basement rocks of the Rangitata Sequence⁴³. The Canterbury Suite-Wakatipu Suite boundary in southern South Island lies within the Haast Schist belt¹, and has not been mapped; the boundary shown here is estimated. In North Island the boundary is placed between the Torlesse and Waipapa terrains¹⁶. Distribution of the ophiolite belt is taken from ref. 53.

67.5% (ref. 10). Tuffs are exceptionally rare. Consistent lithological homogeneity for over 500 km from eastern North Island to Otago (Fig. 1) and in time from Permian to Jurassic indicates the persistence of a continental source capable of supplying a quartzo-feldspathic assemblage for some 100 Myr. Although fossils are generally sparse, those that occur range from late Carboniferous to early Cretaceous^{11,12} and biostratigraphic zonation indicates a westward and northward younging of the sediments^{2,13}.

Based mainly on the distribution of fossil ages¹ deposition of the Canterbury Suite was remarkably episodic with a small pulse in the late Permian (early Tatarian), a very large pulse in the late early to late Triassic (Ladinian–Norian) and a further large pulse in the late Jurassic (early Kimmeridgian to early Tithonian). The extensive occurrence of a thick Jurassic *mélange* belt which includes reworked Triassic Canterbury Suite sediments¹⁴, suggests that much of the Jurassic sedimentation pulse was more related to erosion of the earlier Triassic sediments, as a consequence of uplift associated with the first phase of the Rangitata Orogeny⁸, than to tectonic events in the continental source area.

The distribution of depositional facies, detrital conglomerate, plant fossils and the records of directional sedimentary structures all indicate that sediment transport was in all directions and chiefly to the west and north, with non-marine, marginal marine and shallow marine environments in the east and an ocean basin environment in the west^{1,15}. The Canterbury Suite thus seems to have been a huge clastic wedge⁸ which prograded westward towards a trench, northward to eastern North Island¹⁶, and southward to a peripheral euxenic deep-marine environment at Chatham Island¹⁷.

The stratigraphical and sedimentological evidence points towards an eastern source for the Canterbury Suite. Based on the persistence of a similar composition for ~100 Myr, the quantity of sediment and the rate of sediment supply, a primary cratonic region rather than a late Precambrian or Palaeozoic orogenic belt is indicated. Furthermore, the sandstone composition together with the rate of sediment supply in the Triassic, implies rapid uplift in the source area with vigorous physical erosion and with rapid transport to the sea. In short, there seems to have been an orogeny in the source area which climaxed in the Triassic. Hence, there are mechanistic requirements for the source area in addition to constraints on its composition.

In contrast, the Wakatipu Suite (Fig. 1), derived from a calc-alkaline terrain, shows a pronounced lithostratigraphic change from early and middle Permian volcanics, volcanogenic sediments and intrusives, to late Permian, Triassic and Jurassic highly volcanogenic, mostly marine and locally richly fossiliferous sediments, with common tuffs, but no flows. The complex stratigraphy of the Wakatipu Suite is interpreted^{2,15} as representing the development in the early Permian of a volcanic arc–trench environment followed later by formation of a subduction zone accretionary wedge and mid-slope basin deposits, and finally a regressive frontal arc wedge of upper slope and slope basin, shelf and non-marine sediments.

Jurassic reconstruction of modern Gondwanaland

A central problem in reconstructing the Jurassic configuration is the age of the recurved arc through New Zealand, delineated by the Ophiolite belt (Fig. 1). One idea^{18,19} is that the curve is Cenozoic in age and formed by 500 km of dextral strain associated with movement on the Alpine Fault. Such models show New Zealand as a linear feature in the late Mesozoic^{18,20}, or incorporate 1,000 km of Cenozoic dextral displacement, 500 km of which is attributed to distributed dextral strain^{21–23}. Although some Cenozoic crustal strain cannot be excluded^{24–26}, 500 km of Cenozoic strain requires excessive distortion of the present outline of New Zealand and this has not been substantiated. An alternative idea is that the bending originated mainly in the Rangitata Orogeny (early Jurassic to mid-

Cretaceous⁸)^{27,28}. Palaeomagnetic data from South Island²⁹ and Chatham Island³⁰ (Fig. 1) seem to preclude significant post-late Cretaceous bending between these localities and may support bending during the Rangitata Orogeny. Furthermore, deformation studies³¹ have established the existence of a late Jurassic bend after removing 200 km of Cenozoic dextral strain, an amount which did not excessively distort elements of the land. In the reconstruction to be discussed 200 km of Cenozoic distributed dextral strain is considered to have occurred.

There are two assumptions used in the reconstruction presented as Fig. 2: (1) There has been no major relative movements between Marie Byrd Land and East Antarctica since the late Palaeozoic. (2) There was no relative movement between Marie Byrd Land and the Campbell Plateau in the interval from the late Palaeozoic to the late Cretaceous.

The first assumption is considered to be valid because of the close correspondence in lithologies, ages and structural trends of the Palaeozoic to late Cretaceous geology of North Victoria Land, Marie Byrd Land, New Zealand and Tasmania^{32,33}. Each region was affected by a pre-Silurian–late Ordovician metamorphic event and a widespread phase of late Devonian–Early Carboniferous (350–390 Myr BP) plutonism³³. In Antarctica the trend of the Borchgrevink Orogen extends uninterrupted from the Ford Mountains in western Marie Byrd Land north-northwest through the Ross Sea basement to north-east Victoria Land³⁴. Moreover, two Alban to Neocomian (90–142 Myr BP) calc-alkaline magmatic arcs extended from Queensland through New Zealand to Marie Byrd Land before the late Cretaceous break-up³². In addition, there is no palaeomagnetic evidence to support large displacements (~10³ km) between Marie Byrd Land and East Antarctica after the late Cretaceous–early Cenozoic²², nor is there evidence during this time for the closing of an ocean basin³⁴.

Concerning the second assumption, there are few independent lines of evidence to test the good morphological fit³⁵ of Campbell Plateau and Marie Byrd Land. In pre-Late Cretaceous times Campbell Basin was a continuation of a sediment-filled trough underlying the eastern Ross Sea shelf³⁶. Such a structure would presumably preclude significant late Palaeozoic to late Cretaceous relative movement between Marie Byrd Land and Campbell Plateau.

Because greater New Zealand did not become a landmass until the Rangitata Orogeny (late Triassic to mid-Cretaceous⁸) it is difficult to establish the relative positions of the Canterbury and Wakatipu Suites before the Jurassic, and hence a later Jurassic reconstruction is presented here (Fig. 2). Although some strike-slip faulting may have occurred between the suites in the time between the late early- to late Triassic pulse of sedimentation (Canterbury Suite) and the later Jurassic, it was probably of a small scale (several hundred kilometres), and of dextral character, based on the nature of Cenozoic strike-slip faulting in the New Zealand region.

Source of the Canterbury Suite

The present non-occurrence of a continent east of New Zealand has led to several proposals invoking a source in the west (North Victoria Land–Tasmania–Victoria (Australia)) or in the south-east (Marie Byrd Land). To enable deposition from the west without contamination by volcanoclastic detritus, bypassing mechanisms have been invoked¹³. However, this idea has been largely discredited on the basis of facies properties and sedimentary environments in the Canterbury Suite^{1,37}. Transport between basins of Wakatipu Suite deposition has also been proposed as a working hypothesis for sedimentation and tectonics in the New Zealand geosyncline¹⁶. However, the extent and continuity of the Stokes Magnetic Anomaly System¹⁹ precludes the occurrence of a break in the arc of sufficient size through which the whole of the Canterbury Suite could have been transported without any contamination by volcanoclastic detritus. A development of the western source is the idea that the sediments were derived from an erosional episode in the Carboniferous before the arc–trench environment formed³¹, but

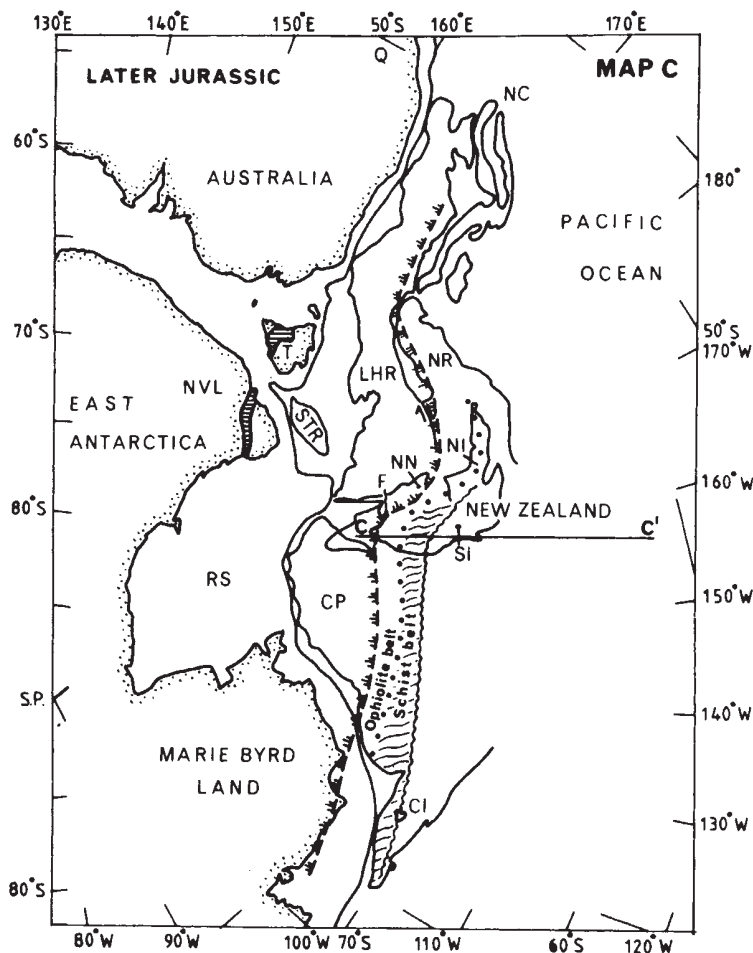


Fig. 2 A later Jurassic palaeogeographic map of eastern Gondwanaland with Antarctica in its present-day position. The New Zealand subcontinent is reconstructed against Norton and Sclater's fit⁵⁴ of Australia and Antarctica. Q, Queensland; NC, New Caledonia; T, Tasmania; NVL, North Victoria Land; LHR, Lord Howe Rise; NR, Norfolk Ridge; STR, South Tasman Rise; F, Fiordland; NN, Northwest Nelson; NI, North Island; SI, South Island; RS, Ross Sea; CP, Campbell Plateau; CI, Chatham Island. The continental outlines of New Zealand⁵⁵, Australia⁵⁶ and Antarctica⁵⁷ are taken as the 2,000-m isobath. The line of 'E' indicates the eastern limit of the pre-late Palaeozoic foreland and the line C-C' relates to the cross-section in Fig. 3. This reconstruction was made by initially establishing the extent, shape and position of the foreland, and then attaching to the foreland margin the 'post-orogenic' form³¹ of the Rangitata Sequence. The southwestern extent of the pre-late Palaeozoic Lord Howe Rise⁵⁸ was taken from borehole data⁵⁹. The margin of the foreland in Campbell Plateau, which has not been mapped, was placed 130 km south of the ophiolite belt⁶³; an equivalent position to the corresponding distance immediately on-land in southeastern South Island (Fig. 1). This position of the margin in Campbell Plateau is consistent with the geology of Campbell Island⁶⁰ (Fig. 1), and is similar to an earlier estimate of the margin⁶², 480 km of Cenozoic^{61,62} dextral transcurrent displacement on the Alpine Fault was reversed. The Waipounami Fault⁶³ is considered to be the boundary off eastern South Island where foreland rocks and the Rangitata sequence of Campbell Plateau have been brought into juxtaposition by Cenozoic drifting²² of the New Zealand sub-continent. The late Cretaceous^{64,65} Bounty Trough and Cenozoic⁶⁶ Waiau Basin-Solander Trough are closed. The fit of Lord Howe Rise against Australia is the conjectural pre-drift position of the Australian margin based on extrapolation of the Australian margin at the time anomaly 33 formed⁶⁷ (82 Myr BP). The reconstruction of Campbell Plateau and Antarctica is a morphological fit³⁵. New Caledonia and Norfolk Ridge are placed against Lord Howe Rise. About 100 km of dextral transcurrent displacement is involved to bring the southern part of Norfolk Ridge into the best visually determined morphological fit with Lord Howe Rise. In this reconstruction North Island occurs in a position slightly further south and oceanward of its present position relative to Lord Howe Rise.

such a theory requires an elaborate mechanism to explain the subsequent wholesale reworking westwards back towards the trench.

An eastern source area is now more acceptable although the location has not been specified^{1,2,15,38}. In attempts to improve the definition of the eastern source it has been proposed³⁷ that the Canterbury Suite was formed from sediment derived from Marie Byrd Land and transported northwestwards. However, the Jurassic reconstruction of New Zealand relative to Gondwanaland presented here (Fig. 2) precludes Marie Byrd Land as the source because it lies on the back-arc side of the Canterbury Suite. Moreover, the shape of the clastic wedge comprising the Canterbury Suite is opposite to that expected if Marie Byrd Land were the source; Chatham Island sediments were deposited in a euxenic deep-marine environment yet in a pre-late Cretaceous position (Fig. 2) occur nearer to Marie Byrd Land than the non-marine and shallow marine deposits more distant in South Island.

An alternative idea also involving Marie Byrd Land as the source, but invoking strike-slip faulting of the Canterbury Suite subparallel to the Palaeo Pacific margin of Antarctica has been proposed^{39,40}, but not applied to a rigorous Mesozoic reconstruction of eastern Gondwanaland and New Zealand. In taking the later Triassic reconstruction presented here (Fig. 3), and noting particularly the extent of the volcanic arc, to avoid transportation through the arc-trench system some 2,000 km of sinistral strike-slip movement would be required to derive the Canterbury Suite from more eastern parts of Marie Byrd Land. Moreover, the bulk of this movement would need to have occurred between the late early and late Triassic pulse of sedimentation and the first phase of the Rangitata Orogeny (late Triassic-early Jurassic⁸). No mechanism has been proposed and substantiated to account for such rapid movement, and it is in an

opposite sense to Cenozoic strike-slip movement in New Zealand. A model invoking strike-slip faulting would be acceptable only if a lesser amount of fault movement were necessary. This could apparently only be achieved by arbitrarily twisting the Campbell Plateau and Lord Howe Rise into different positions during the Mesozoic, such that until the mid-Jurassic the Canterbury Suite was more proximal to source, yet, by the mid-Cretaceous the Campbell Plateau and Lord Howe Rise were in the pre-breakup late Cretaceous positions. There is no evidence to suggest that before the mid-Cretaceous the Campbell Plateau and Lord Howe Rise had different positions relative to Antarctica and Australia, and hence there is not a substantiated case for Marie Byrd Land as the source of the Canterbury Suite.

Pacifica and development of New Zealand

Pacifica was a continent hypothesized from geological evidence³ to have lain off eastern Gondwanaland until the Triassic. It then fragmented and now occurs as exotic blocks of continental crust within the circum-Pacific Cordillera. The recent demonstration⁴¹ that large parts of western North America are allochthonous adds credibility to Pacifica's fragmenting and subsequently contributing to formation of the Pacific Cordillera.

In discussing Lost Pacifica, Nur and Ben Avraham³ did not use stratigraphical evidence from New Zealand to support their argument. Pacifica may well be the source of the Canterbury Suite⁴². Not only was Pacifica in the correct position at the right time, but its fragmentation also provides the mechanism for deriving huge volumes of Permian and Triassic sandstone.

I largely agree with the plate tectonic interpretations^{2,4,31} for the Wakatipu Suite and the subsequent collision of the two suites, but seek to extend these models eastwards by including Pacifica (Fig. 3).

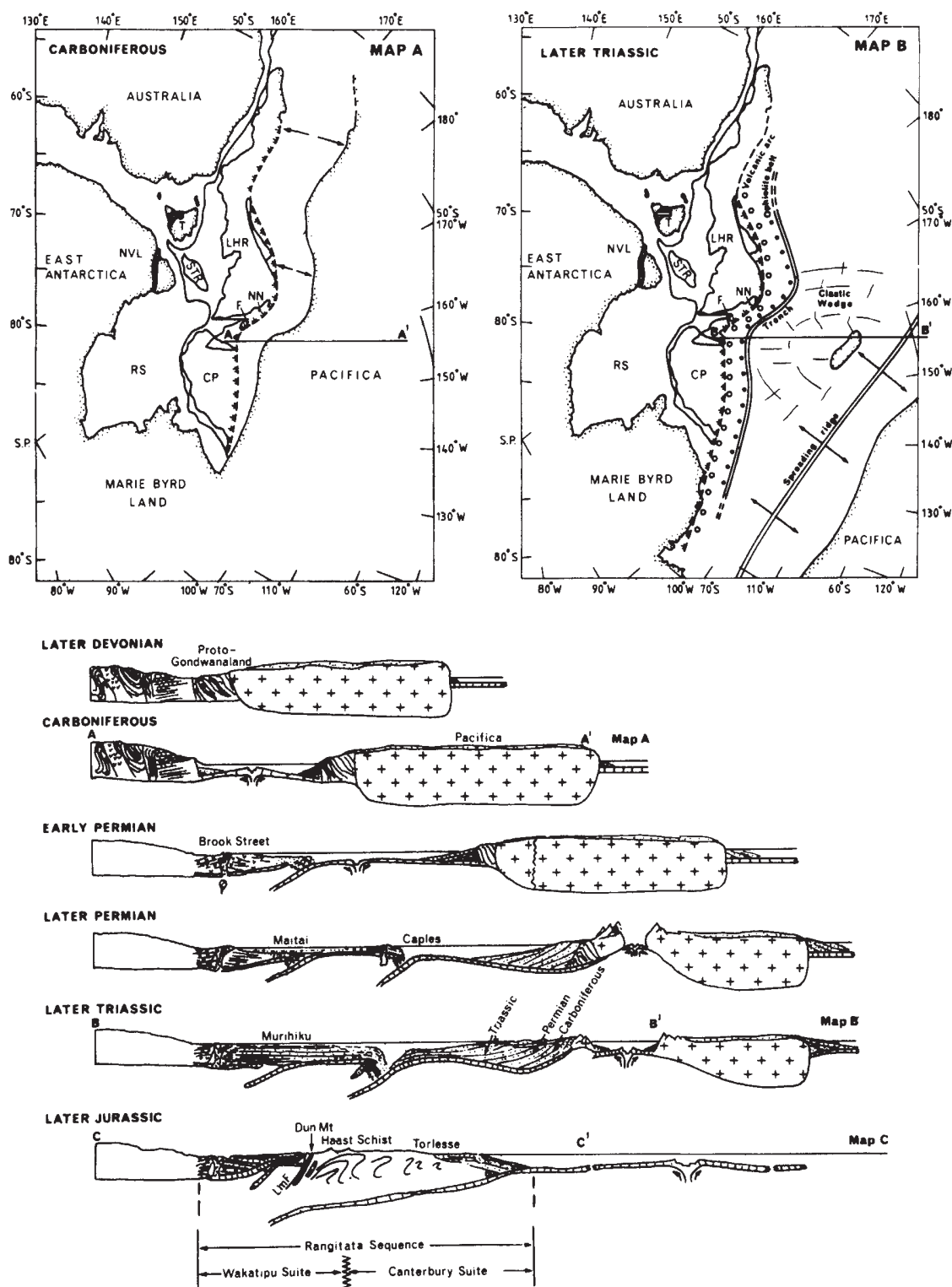


Fig. 3 Palaeogeographic maps and schematic cross-sections illustrating a model of the late Palaeozoic-Mesozoic geological development of New Zealand. These show a relationship of the plate tectonic development of the Wakatipu Suite^{2,31} and derivation of the Canterbury Suite, to the drift and fragmentation of Pacifica. The later Jurassic cross-section relates to the section line in Fig. 2. The details of map construction and abbreviations are given in Fig. 2, and the terrain names on the cross-sections (for example, Brook Street) are defined elsewhere².

The Rangitata sedimentary-orogenic cycle⁴³ is postulated to have begun in the late Devonian to early Carboniferous with rifting within the Tuhua Orogen (Fig. 3); fragmentation of supercontinents is believed commonly to occur along old orogenic belts⁴⁴. The spreading centre probably propagated southwards. A reconstruction of New Zealand for the Carboniferous (Fig. 3, map A) shows that the foreland was curved and

the corresponding embayment in Pacifica might have promoted delta formation of the Canterbury Suite. By the Early Permian the Wakatipu arc-trench environment had formed and by the mid-Permian an incipient fracture had developed within Pacifica. The fragmenting of Pacifica (Fig. 3) ensured extensive uplift of the continent during the late Permian and Triassic and derivation, by physical erosion, of detritus which preserved a

quartzo-feldspathic composition because of the huge volumes involved and the rapid transport to the sea. The spreading ridge which initiated the fragmentation of Pacifica then extended south to Antarctica (Fig. 3, map B), perhaps associated with anticlockwise rotation of Gondwanaland concomitant with transformation from Panagea B to A⁴⁵, and by the end of the Triassic was actively rafting the Canterbury Suite on oceanic crust towards the Wakatipu Suite. The first phase of the Rangitata Orogeny ensued in the late Triassic–early Jurassic⁸, and although most folding occurred in the Canterbury Suite, the wedge impinged on the Wakatipu Suite, consequently restricting Jurassic sedimentation to the areas now represented by the Southland and Kawhia Synclines (Fig. 1). The point collision of the clastic wedge (Fig. 3, map B) caused eversion of the Ophiolite and its progressive convergence westwards with the foreland (Fig. 2). In the Jurassic, Pacifica fragmented further and parts were drifting across the Pacific Ocean.

Source of the Franciscan assemblage

Attention has been drawn previously to the geological similarity between California and New Zealand: each comprises an inner volcanogenic terrain (Great Valley sequence versus Wakatipu Suite) adjacent to an older crystalline foreland, and an outer rarely fossiliferous quartzo-feldspathic terrain (Franciscan assemblage versus Canterbury Suite)⁴. As in New Zealand, the critical problem is the source of the quartzo-feldspathic assemblage which also has a composition close to granodiorite, and was deposited on the Pacific side of the coeval Great Valley sequence. Although the Franciscan assemblage is derived from a continental source, the differences in grain-size and composition between the two sedimentary belts, and the pronounced westward thinning of the Great Valley sequence, are difficult to reconcile with a Sierran–Klamath source for the Franciscan assemblage⁴. This difficulty led to the proposal that a remnant volcanoplutonic arc, possibly a continental fragment, existed between the two sedimentary terrains during the late Jurassic to early Tertiary⁴⁶. Subsequently, the remnant arc was eroded to derive the Franciscan assemblage and its roots subducted so that

now no part remains⁴. This remnant arc could well have been a fragment of Pacifica, for such a fragment approached California during the late Jurassic to early Cretaceous³. Consistent with this model, the bulk of the Franciscan assemblage was deposited in latest Jurassic to earliest Cretaceous time (Tithonian–Valanginian)⁴, when presumably the continental fragment was rapidly uplifted and eroded due to interaction with the North American Plate.

Consequences of Gondwanaland rifting

Pacifica is conceived as a large continent not only adjacent to New Zealand but also to Antarctica³. The age and distribution of Jurassic tholeiitic dykes and sills and basaltic lavas from Africa across Antarctica to Tasmania show evidence of a period of incipient rifting within Gondwanaland well before the separation of East and West Gondwanaland at 140–160 Myr BP. In Africa these igneous rocks are the Karroo Basalts dated⁵ at 150–190 Myr, in Antarctica the 175±5 Myr old Dufek Massif, Ferrar Dolerites and Kirkpatrick Basalts⁴⁷, and in Tasmania, dolerites aged⁷ at 170.5±5 Myr. Rather than link these voluminous rocks to the break-up of East and West Antarctica⁴⁸, an abortive separation of India⁴⁹, an initial rifting of East and West Gondwanaland⁵⁰, or incipient rifting of Australia and Antarctica⁵¹, for all of which the age of rifting post-dates the intrusion and extrusion of these rocks, their origin may be related to contemporary volcanicity and active spreading in the Palaeo Pacific Ocean which caused the breakup and drift of Pacifica away from the Pacific margin of Gondwanaland. The occurrence of the African, Antarctic and Tasmanian Jurassic dolerites and basalts has already been suggested as related to a former coeval spreading axis in the Palaeo Pacific Ocean off Antarctica⁴⁴.

I thank Dr C. S. Nelson, Professor M. J. Selby and Dr A. P. W. Hodder for encouragement and discussion, and Dr J. D. Bradshaw for discussion and access to unpublished work, especially relating to the nature and timing of the Rangitata Orogeny.

Received 5 August; accepted 22 October 1980.

- Andrews, P. B., Speden, I. G. & Bradshaw, J. D. *N.Z. J. Geol. Geophys.* **19**, 791–819 (1976).
- Coombs, D. S. *et al. Am. J. Sci.* **276**, 561–603 (1976).
- Nur, A. & Ben Avraham, Z. *Nature* **270**, 41–43 (1977).
- Blake, M. C., Jones, D. L. & Landis, C. A. in *The Geology of Continental Margins* (eds Burk, C. A. & Drake, C. L.) 853–872 (Springer, New York, 1974).
- Cox, K. G. in *African Magmatism and Tectonics* (eds Clifford, T. N. & Gass, I. G.) 211–236 (Oliver and Boyd, Edinburgh, 1970).
- Barrett, P. J., Grindley, G. W. & Webb, P. N. in *Antarctic Geology and Geophysics* (ed. Adie, R. J.) 319–332 (Oslo University Press, 1972).
- Schmidt, P. W. & McDougall, I. *J. geol. Soc. Aust.* **25**, 321–328 (1977).
- Bradshaw, J. D., Andrews, P. B. & Adams, C. J. *5th Gondwana Symp. Victoria University of Wellington Abstr.* (IUGS Subcommittee on Gondwana Stratigraphy and Paleontology, Wellington, 1980).
- Mutch, A. R. *Trans. R. Soc. N.Z.* **84**, 499–511 (1957).
- Reed, J. J. *Bull. geol. Surv. N.Z.* **57**, 1–60 (1957).
- Campbell, J. D. & Warren, G. *Trans. R. Soc. N.Z., Geol.* **3**, 99–137 (1965).
- Jenkins, T. B. H. & Jenkins, D. G. *Nature* **233**, 117–118 (1971).
- Landis, C. A. & Bishop, D. G. *Bull. geol. Soc. Am.* **83**, 2267–2284 (1972).
- Bradshaw, J. D. *R. Soc. N.Z. J.* **3**, 161–167 (1973).
- Carter, R. M., Hicks, M. D., Norris, R. J. & Turnbull, I. M. in *Sedimentation in Submarine Canyons, Fans and Trenches* (eds Stanley, D. J. & Kelling, G.) 340–361 (Dowden, Hutchinson and Ross, Stroudsburg, 1978).
- Spörli, K. B. *Bull. geol. Soc. Am.* **89**, 415–425 (1978).
- Adams, C. J. D. & Robinson, P. N. *N.Z. J. Geol. Geophys.* **20**, 287–301 (1977).
- Wellman, H. W. *Bull. Aust. Soc. Explor. Geophys.* **6**, 55–56 (1975).
- Hunt, T. *N.Z. J. Geol. Geophys.* **21**, 596–606 (1978).
- Cooper, R. A. *Geol. Soc. N.Z. Newslett.* **42**, 31–34 (1976).
- Molnar, P., Atwater, T., Mamerickx, J. & Smith, S. M. *Geophys. J. R. astr. Soc.* **40**, 383–420 (1975).
- Weissel, J. K., Hayes, D. E. & Herron, E. M. *Mar. Geol.* **25**, 231–277 (1977).
- Crook, K. A. W. & Belbin, L. *J. geol. Soc. Aust.* **25**, 23–40 (1978).
- Lensen, G. J. *Bull. geol. Soc. Am.* **79**, 545–56 (1968).
- Walcott, R. I. *Geophys. J. R. astr. Soc.* **52**, 137–64 (1978).
- Norris, R. J. in *The Origin of the Southern Alps*, Bull. 18 (eds Walcott, R. I. & Cresswell, M. M.) (The Royal Society of New Zealand, Wellington, 1979).
- Fleming, C. A. *Q. J. geol. Soc. Lond.* **125**, 125–170 (1970).
- Waterhouse, J. B. *Pacif. Geol.* **9**, 35–73 (1975).
- Oliver, P. J., Mumme, T. C., Grindley, G. W. & Vella, P. *N.Z. J. Geol. Geophys.* **22**, 119–212 (1979).
- Grindley, G. W., Adams, C. J. D., Lumb, J. T. & Watters, W. A. *N.Z. J. Geol. Geophys.* **20**, 425–467 (1977).
- Wood, B. L. *Tectonophysics* **47**, 339–368 (1978).
- Harrington, H. J. *J. geol. Soc. Aust.* **26**, 276–277 (1979).
- Adams, C. J. *5th Gondwana Symp. Victoria Univ. of Wellington Abstr.* (IUGS Subcommittee on Gondwana Stratigraphy and Paleontology, Wellington, 1980).
- Elliott, D. H. *Am. J. Sci.* **275A**, 45–106 (1975).
- Griffiths, J. R. *Nature* **249**, 336–338 (1974).
- Davey, F. J. & Houtz, R. E. *Mar. Geol.* **25**, 61–72 (1977).
- Bradshaw, J. D. & Andrews, P. B. *Nature* **241**, 14–16 (1973).
- Retallack, G. J. *R. Soc. N.Z. J.* **9**, 397–414 (1979).
- Norris, R. J., Landis, C. A. & Ward, C. M. *Geol. Soc. N.Z. Queenstown Conf. Abstr.* (1977).
- Ward, C. M. & Spörli, K. B. *J. Geol.* **87**, 187–193 (1979).
- Monger, J. W. H. & Irving, E. *Nature* **285**, 289–294 (1980).
- Pirajno, F. *Geol. Soc. N.Z. Newslett.* **49**, 19–21 (1980).
- Carter, R. M., Landis, C. A., Norris, R. J. & Bishop, D. G. *R. Soc. N.Z. J.* **4**, 5–18 (1974).
- Sutton, J. *Phil. Trans. R. Soc. B279*, 197–205 (1977).
- Irving, E. *Nature* **270**, 304–309 (1977).
- Blake, M. C. & Jones, D. L. *Soc. Econ. Paleontol. Miner. Spec. Publ.* **19**, 255–263 (1974).
- Kyle, P. R., Elliott, D. H. & Ford, A. B. *5th Gondwana Symp. Victoria Univ. Wellington Abstr.* (IUGS Subcommittee on Gondwana Stratigraphy and Paleontology, Wellington, 1980).
- Tarling, D. H. *Nature* **238**, 92–93 (1972).
- Scrutton, R. A. *Geol. Mag.* **110**, 227–234 (1973).
- Barron, E. J., Harrison, C. G. A. & Hay, W. W. *EOS* **59**, 436–449 (1978).
- Veevers, J. J. & Evans, P. R. in *Gondwana Geology* (ed. Campbell, K. S. W.) 579–607 (ANU Canberra, 1975).
- Cooper, R. A. *R. Soc. N.Z. J.* **9**, 29–84 (1979).
- Davey, F. D. & Christoffel, D. A. *Earth planet. Sci. Lett.* **41**, 14–20 (1978).
- Norton, I. D. & Sclater, J. G. *J. geophys. Res.* **84**, 6803–6830 (1979).
- N.Z. Oceanographic Institute *Tectonic Map of the South West Pacific* (DSIR, Wellington, 1970).
- Circum Pacific Council for Energy and Mineral Resources, *Geographic Map of the Circum-Pacific Region* (American Association of Petroleum Geologists, Tulsa, 1978).
- National Mapping Office, Dept. of the Interior, *US Antarctica* (Map, 2nd edn) (Antarctic Division, Dept of External Affairs, US, 1956).
- McDougall, I. & van der Linde, G. J. *Earth planet. Sci. Lett.* **21**, 117–126 (1974).
- Wodzicki, A. *N.Z. J. Geol. Geophys.* **17**, 747–757 (1974).
- Fleming, C. A. *Tuatara* **10**, 52–108 (1962).
- Carter, R. M. & Norris, R. J. *Earth planet. Sci. Lett.* **31**, 85–94 (1976).
- Cutten, H. N. C. *N.Z. J. Geol. Geophys.* **22**, 535–553 (1979).
- Cullen, D. J. *N.Z. J. mar. Freshw. Res.* **1**, 16–25 (1967).
- Houtz, R. E., Ewing, J., Ewing, M. & Lonard, A. G. *J. geophys. Res.* **72**, 4713–4739 (1967).
- Kennett, J. P. & Casey, R. E. *N.Z. J. mar. Freshw. Res.* **3**, 20–28 (1969).
- Norris, R. J., Carter, R. M. & Turnbull, I. M. *J. Geol. Soc. Lond.* **135**, 119–205 (1978).
- Weissel, J. K. & Hayes, D. E. *Earth planet. Sci. Lett.* **36**, 77–84 (1977).

Moloney murine sarcoma proviral DNA is a transcriptional unit

Edmund W. Benz Jr*, Robert M. Wydro†, Bernardo Nadal-Ginard† & Dino Dina‡

* Department of Developmental Biology and Cancer, Albert Einstein College of Medicine, Bronx, New York 10461

† Department of Cell Biology, Albert Einstein College of Medicine, Bronx, New York 10461

‡ Department of Genetics and Department of Microbiology and Immunology, Albert Einstein College of Medicine, Bronx, New York 10461

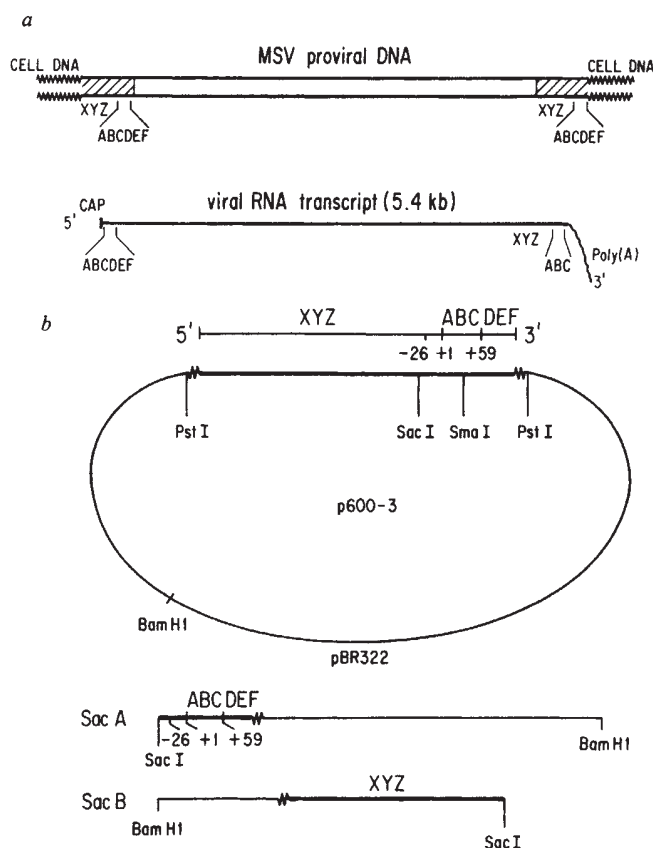
A portion of Moloney murine sarcoma virus DNA which is repeated at both ends of the provirus has been sequenced. The nucleotide sequence, together with hybridization data obtained with in vitro pulse-labelled nascent viral RNA, indicate that initiation and termination of RNA synthesis occur within that region of the proviral DNA. A model for transcriptional readthrough of termination signals during RNA synthesis in this system is suggested.

A CRUCIAL STEP in the life cycle of retroviruses is their integration as double-stranded DNA molecules into the DNA of the cell they infect (for a review see ref. 1). The mechanism(s) which regulates transcription of integrated DNA into viral RNA and viral mRNAs has been actively investigated. However, although it is known that cellular RNA polymerase II mediates transcription of the viral genome²⁻⁵, little is known about the actual sites for initiation and termination of transcription; for example, whether these sites reside on host cell DNA or viral DNA. The size of nuclear virus-specific primary transcripts has not been established; data supporting the presence of both genome-length and longer than genome-length nuclear precursors⁶⁻⁹ have recently been reported. The discovery that integrated and unintegrated viral DNAs contain a direct repeat at both ends of the DNA and thus are longer than the viral RNA template¹⁰⁻¹⁵ has raised some interesting questions about the potential of the proviral DNA to function as a transcriptional unit. In the case of Moloney murine sarcoma virus (MSV), the

viral DNA carries a 600, base-pair long terminal direct repeat (LTR), which contains the sequences termed XYZABCDEF in Fig. 1a¹¹. The mature viral RNA starts and terminates within the LTR with the specific sequence termed 'ABC' (Fig. 1a). This sequence is 59 nucleotides long and is present at both ends of the viral RNA. Its role in reverse transcription has recently been elucidated^{16,17}. The genome structure shown in Fig. 1a implies that if MSV functions as a transcriptional unit, the LTR must contain sequences that code for the initiation and termination of RNA synthesis. Alternatively, if transcription starts and terminates outside the proviral DNA, the LTR must carry the information necessary for RNA processing.

We report here DNA sequence data obtained from the LTR DNA cloned into a bacterial plasmid, pBR322. The nucleotide sequence upstream from the region of DNA which codes for the 5' end of the viral RNA (capping site) contains the sequence CAATAAAA, a variation of the typical 'Hogness box', which has been associated with promotion or initiation of transcription

Fig. 1 a, Relationship between genomic MSV RNA and integrated DNA. The structure of MSV DNA has been drawn according to results obtained with unintegrated and integrated MSV double-stranded DNA^{11,45,46}. The 600-base pair direct repeats (LTRs) are represented by hatched areas. The sequence ABC corresponds to the 59-nucleotide repeat present at either end of viral RNA (ref. 32 and Fig. 3); transcription is from left to right. kb, Kilobases. **b**, Schematic representation of recombinant plasmid p-600-3 from which SacA and SacB DNA fragments were derived. The 600-base pair DNA (LTR) was prepared *in vitro* and purified on agarose gels as described previously¹⁷. Briefly, S₁ nuclease-treated LTR DNA was tailed with oligo(dC) and plasmid pBR322, cleaved at the PstI site and tailed with oligo(dG). After annealing, the recombinant molecules were selected and amplified in *E. coli* χ 1776. Screening was performed with 'strong-stop' MSV cDNA^{17,47}. The map of pBR322 is drawn in an anticlockwise orientation. The insert represents the majority of the LTR DNA present on full-length genomic DNA. Its length is 572 nucleotides, of which 518 are viral and the rest G-C tails. Fifty nucleotides are missing from the right end of the sequence and ~15 nucleotides have been deleted at the left end. The numbers +1, +59 and -26 indicate the coordinates of RNA transcription (see Fig. 2) which proceeds from left to right. The SacA and SacB fragments were obtained by digestions with the restriction endonucleases SacI and BamHI. DNA fragments were separated on 5% polyacrylamide gels and electroeluted into dialysis bags. After ethanol precipitation, 1 μ g of purified DNA was loaded on to Millipore filters. The filters were processed for hybridization and stored under vacuum.



in many of the eukaryotic genes which have been sequenced. The Hogness box is followed by a AATAAA poly(A)-addition signal 70 nucleotides downstream. Hybridization data obtained with *in vitro* pulse-labelled nuclear RNAs indicate that initiation of transcription occurs close to or at sequences corresponding to the viral RNA capping site. The relative position of initiation and polyadenylation sites suggests that a readthrough mechanism may be used in synthesis of mature viral RNA. The potential secondary structure of the RNA transcribed from this portion of the viral genome suggests that in MSV, chain termination may occur by a mechanism similar to attenuation in bacteria^{18,19}. A model for such a regulatory mechanism is suggested here.

DNA sequence data

The long terminal repeat (LTR) present on the unintegrated and integrated viral DNA participates in several crucial steps of the virus life cycle, including DNA synthesis, integration and transcription. To establish whether any specific structural features contained within the LTR might clarify its role in transcription, we have determined the nucleotide sequence of a recombinant plasmid (p-600-3) carrying an insert which represents 518 base pairs of the LTR DNA (Fig. 1b).

Figure 2 shows the nucleotide sequence of the last 177 base pairs (3' end) of the LTR DNA. A Hogness box, which has been found in the majority of eukaryotic genes transcribed by RNA polymerase II (refs 20-23 and ref. 24 for a review), is found 26 nucleotides upstream from the capping site. Although this sequence has been previously considered the structural equivalent of a bacterial 'Pribnow box' (ref. 20), and therefore a promoter, recent evidence suggests that it confers homogeneity

to the starting nucleotide or capping site of the transcripts²³. In its absence, heterogeneous 5' ends are found, but initiation of RNA synthesis is maintained^{22,23}. The presence of such a structural element on MSV DNA argues in favour of the possibility that synthesis of mature viral RNA may actually start at this site and that the nucleotide sequence surrounding the capping site and including the Hogness box may serve as a promoter *in vivo*.

A second specific structure, reported for many of the genes transcribed by RNA polymerase II, is the polyadenylation signal AATAAA. This sequence, found at position +45, is followed by another sequence, TTGC or TTTGC^{24,25}, at position +63 to +67, immediately downstream from the presumed terminal tetranucleotide TTGC. Comparative data for several other mRNAs²⁴ and viral RNAs^{25,26} indicate that this region is associated with termination of RNA transcription. If transcriptional processes *in vivo* actually use the structural regions described above as regulatory signals, it follows that the LTRs of integrated proviral DNA have a promoter region immediately adjacent to terminator signals.

Does transcription of MSV RNA start at a cellular or viral promoter?

The presence of sequences in the LTR DNA which correspond to the putative promoter sites of several eukaryotic genes²⁴ transcribed by RNA polymerase II raises the specific question of whether or not virion RNA is the primary product of nuclear transcription in infected cells. We have therefore isolated nascent, pulse-labelled nuclear RNA synthesized *in vitro* and analysed it by hybridization to different portions of the p-600-3 plasmid DNA (Fig. 2). If transcription started at a cellular site upstream from the viral 'promoter site' (marked -26 in Fig. 1b), LTR sequences (XYZ) upstream from the *SacI* site would be transcribed and covalently linked to downstream sequences (ABCDEF). If instead transcription started at the putative viral promoter site (Fig. 1b), the 5' ends of virus-specific nuclear RNAs would correspond only to sequences downstream from the *SacI* site (ABCDEF). By hybridizing *in vitro* pulse-labelled nuclear RNA to LTR sequences downstream from the promoter (*SacA* fragment, Fig. 1b) or LTR sequences upstream from the promoter (*SacB* fragment, Fig. 1b), the relative amounts transcribed from these two regions of the LTR DNA have been quantified. This analysis must, however, take into account the fact that mature virion RNA ends with the sequence XYZABC, which can hybridize to both the *SacA* and *SacB* fragments. (Table 1 summarizes the distribution of virus-specific nucleotides on the *SacA* and *SacB* fragments.) Therefore, before hybridization, we removed polyadenylated mature viral RNA from the nuclear RNA sample by chromatography on oligo(dT)-cellulose (see Table 2 legend). Mature virion RNA was fragmented by alkaline treatment²⁷ and fractionated by oligo(dT) chromatography. Viral poly(A)⁺ and poly(A)⁻ RNA fragments were used as controls in the hybridization to DNA-nitrocellulose filters. DNA-nitrocellulose filters carrying specific DNA fragments were prepared as described in Fig. 2 legend.

```

      -30      -20      -10      +1      +10
      AGCTCAA TAAAGAGCC CACAACCCCT CACTCGGCGC GCCAGTCTTC
      TCGAGTT ATTTCTTCGG GTGTTGGGGA GTGAGCCGCG CGGTCAAGAAG

      +20      +30      +40      +50      +60
      CGATAGACTG CGTCGCCCGG GTACCCGTAT TCCCAATAAA GCCTCTTGCT
      GCTATCTGAC GCAGCGGGCC CATGGGCATA AGGTTATTTT CGGAGAACGA

      +70      +80      +90      +100     +110
      GTTTCATCC GAATCGTGGT CTCGCTGTTT CTTGGGAGGG TCTCCTCTGA
      CAAACGTAGG CTTAGCACCA GAGCGACAAG GAACCTCCCG AGAGGAGACT

      +120     +130     +140
      GTGATTGACT ACCCAGCAGC GGGGTCTTTC A
      CACTAAGTGA TGGGTGCTGC CCCCAGAAAG T
  
```

Fig. 2 Sequence of the right end of the LTR DNA. For sequencing, the 572-base pair insert was released from the plasmid by digestion with *PstI* and purified on a 5% polyacrylamide gel (acrylamide/bis = 30:1, 20 cm × 20 cm × 1.0 cm). This fragment was digested with *PvuII*, *AvaI*, *HhaI*, *SmaI* or *SacI* and end labelled with T4 polynucleotide kinase and [γ -³²P]ATP (3,000 Ci mmol⁻¹, Amersham) or terminal transferase and 3'-[α -³²P]dATP (1,000 Ci mmol⁻¹, NEN). For strand separation the labelled DNA fragments were dissolved in 30 μ l of 0.3 M NaOH, 10% glycerol (v/v), 1 mM EDTA and 0.05% bromophenol blue and xylene cyanol, heated at 90 °C for 2 min, chilled on ice and immediately fractionated on a 5% polyacrylamide (acrylamide/bis, 60:1) strand-separation gel at 300 V at room temperature in 50 mM Tris-borate, pH 8.3, 1 mM EDTA buffer. These singly end-labelled fragments were sequenced by the chemical cleavage method of Maxam and Gilbert⁴⁸. The nucleotide sequence of both DNA strands was determined except for the last 50 nucleotides. This portion of the sequence has been obtained from strong-stop MSV cDNA synthesized *in vitro*¹⁷. +1 Indicates the first nucleotide on the viral RNA map, +59 the last one. The Hogness box and the polyadenylation signals are underlined.

Table 1 Distribution of virus-specific sequences on *SacA* and *SacB* fragments

Fragment	Total virus-specific nucleotides	%	Nucleotides complementary to 5' viral RNA		Nucleotides complementary to 3' viral RNA	
				%		%
p-600-3	518	100	91	17	486	94
<i>SacA</i>	128	25	91	17	96	18
<i>SacB</i>	390	75	0	0	390	75

The number of nucleotides on each fragment was derived from the sequencing data presented in Fig. 2 and unpublished results. The number of hybridizable virus-specific nucleotides from either end of the viral RNA was calculated by taking into account the presence of the 59-nucleotide ABC repeat at the end of the viral RNA. The *SacA* fragment contains 'promoter-downstream' sequences (ABCDEF) and the *SacB* fragment carries 'promoter-upstream' sequences (XYZ) (Fig. 1b).

Table 2 Hybridization of pulse-labelled nuclear RNA from MSV-infected cells to p-600 DNA fragments

RNA sample	Input c.p.m.	c.p.m. hybridized to SacB (%)	c.p.m. hybridized to SacA (%)
Nuclear poly(A) ⁺	1.8 × 10 ⁶	373 (75)	123 (25)
Viral poly(A) ⁺	3.2 × 10 ⁴	1,116 (72)	446 (28)
Nuclear poly(A) ⁻	1.2 × 10 ⁷	910 (40)	1,365 (60)
Viral poly(A) ⁻	1.12 × 10 ⁵	2,507 (30)	5,988 (70)

Approximately 2×10^7 cells were directly lysed on the plate by addition of lysis buffer (25 mM Tris-HCl, pH 7.4, 25 mM NaCl, 5 mM MgCl₂, 140 mM sucrose, 0.5% Triton X-100). After 10 min in the cold, cells were scraped off the plates with a Teflon-coated razor blade. Complete lysis was achieved by subjecting the cells to 10 strokes with a B pestle in a Dounce homogenizer. Nuclei were pelleted through a sucrose cushion (0.25 M sucrose, 5 mM MgCl₂, 25 mM Tris pH 7.4) at 4 °C with a table-top centrifuge. The nuclear pellet was resuspended in 1.5 ml of buffer (0.25 M sucrose, 5 mM MgCl₂, 25 mM Tris pH 7.4) and pelleted again. Final recovery was 1.1×10^7 nuclei. The crude nuclei were gently suspended in 150 µl of a solution containing 125 mM Tris pH 7.9, 2 mM MgCl₂, 5 mM NaF, 1 mg ml⁻¹ bovine serum albumin, 1 mM ATP, GTP and CTP, 5.8 µM [α -³²P]UTP, 50 mM (NH₄)₂SO₄ and 1 mM dithiothreitol. The reaction was incubated at 37 °C for 5 min and terminated by the addition of 500 µl of 50 mM Na-acetate pH 5, 6 mM EDTA and 1% SDS. Total incorporation measured by trichloroacetic acid precipitation was $\sim 3 \times 10^7$ c.p.m. Nuclear RNA was extracted several times with phenol saturated with 0.1 M Na-acetate pH 5.0. After ethanol precipitation, the poly(A)-containing nuclear RNA was fractionated by chromatography on oligo(dT)-cellulose. An aliquot of each RNA fraction was then hybridized to specific p-600 DNA fragments bound to nitrocellulose filters in 1 ml of HEPES hybridization buffer (100 mM HEPES, pH 7, 1 mM EDTA, 0.2% SDS, 0.5 M NaCl and 50% formamide, 48 h at 40 °C). The filters were washed extensively in 2 × SSC at 65% and treated with RNase (20 µg ml⁻¹, 2 × SSC at room temperature for 1 h). Hybridization experiments were performed in triplicate. In these experiments nonspecific background values (10 c.p.m. per 10⁶ c.p.m.) were subtracted from the results presented above.

The nuclear RNA samples labelled with [α -³²P]UTP and mature virion RNAs uniformly labelled with ³²P-orthophosphoric acid were hybridized to DNA filters stacked in a scintillation vial (Table 2). If the oligo(dT)-cellulose chromatography step removed all the RNA chains containing the XYZABC portion of the LTR located at the 3' end RNA, the viral poly(A)⁻ RNA fragments should hybridize exclusively to the SacA DNA-containing filter. Although the majority of poly(A)⁻ RNA counts hybridized to SacA DNA (Table 2), a substantial background (30%) of viral poly(A)⁻ RNA hybridized to SacB DNA, implying that oligo(dT)-cellulose chromatography could not completely remove all 3'-terminal fragments of viral RNA molecules. This was not surprising as 10–30% of full-length virion RNA lacks a poly(A) tail (ref. 28 and D.D., unpublished observation). Similarly, we observed that 60% of virus-specific, poly(A)⁻, pulse-labelled nuclear RNA counts hybridized to SacA DNA whereas 40% hybridized to SacB DNA (Table 2). This result implied that pulse-labelled, virus-specific nuclear RNA was very similar to mature virion RNA and that newly synthesized nuclear RNA started primarily at a site on the SacA fragment.

To eliminate the ambiguity due to the background hybridization of 3' sequences in the experiment described above, another experiment was devised. The pulse labelling was performed in the presence of ribonucleoside triphosphate analogue which contained a sulphur atom substituted for oxygen in the γ phosphate position (γ -S NTPs) (see Table 3 legend). Newly initiated chains have the structure γ -S pppN ³²pUp. The incorporation of a γ -S nucleotide at the 5' end of a chain prevents capping of that chain (Furichi and R. N. Wydro, unpublished results), thus allowing its purification by mercury-agarose affinity chromatography²⁹. Therefore, RNAs retained by the column terminate with 5' γ -S triphosphate, indicating that these RNAs result from initiation during the pulse label and not from elongation. Because most RNA chains in eukaryotic nuclei start with G or A, only those two substrates were actually used.

After fractionation of the RNA on mercury-agarose columns (see Table 3), a fraction containing the bulk of the nuclear RNA (unbound RNA) and a fraction containing γ -S pppA- and γ -S pppG-initiated chains were obtained (Table 3). The percentage of RNA chains initiated with a substituted nucleotide was

approximately 3% of the total [α -³²P]UTP-labelled chains, in agreement with previously obtained results^{30,31}. The 'bound' and 'unbound' fractions were hybridized to nitrocellulose filters carrying the SacA and SacB fragments as described above. The hybridization pattern of the unbound fractions of the RNA made in the presence of γ -S ATP and γ -S GTP were quantitatively similar, being proportional to the number of virus-specific nucleotides present on the SacA and SacB fragments (Table 1). This result showed that the substituted nucleotides did not interfere significantly with the elongation of virus-specific RNA chains and that transcription of the whole viral genome occurred. γ -S pppG-initiated chains hybridized preferentially to the SacA fragment. When normalized to the number of virus-specific nucleotides of the SacA and SacB fragments (Table 1), the ratio of the counts hybridized to SacA compared with SacB DNA was greater than 20:1. From these data we concluded that most of the virus-specific RNA chains were initiated within the SacA fragment. The counts which hybridized to the SacB fragment could be due either to the presence of contaminating total RNA in the bound fraction or to a small number of chains initiated at a cellular promoter. The small number of hybridized counts obtained with γ -S pppA-initiated chains supports the conclusion that sulphur groups were not transferred to internal residues of elongating chains.

A model for chain termination

The evidence presented here strongly suggests that initiation of viral RNA synthesis occurs at a site on proviral DNA contained within the last 177 nucleotides of the 600-base pair repeat. Because the 5' and 3' ends of the primary transcripts of many RNA polymerase II-dependent transcriptional units are conserved during processing of precursor nuclear RNA into mature mRNA (for a summary see refs 24, 25), it is reasonable to assume that virus-specific nuclear RNA chains have the same 3' end as mature virion RNA. As shown in Fig. 3, we have tentatively placed the nucleotide which precedes the poly(A) at position +59 on the basis of oligonucleotide sequence data (ref. 32 and unpublished observations). The DNA sequence between the capping site (+1) and the region immediately following the end of the viral RNA (+59) thus contains signals for poly(A) addition and possibly chain termination. These signals must be ignored by RNA polymerase II during transcription of the LTR at the left end of the proviral DNA (see Fig. 1a); however, the enzyme would have to recognize such signals at the end of transcription, 5,400 nucleotides downstream, as it transcribes the LTR at the right end of the genome.

Table 3 Hybridization of RNA from MSV-infected cell nuclei pulse-labelled in the presence of γ -S GTP and γ -S ATP to p-600-3 DNA fragments

RNA sample	Input c.p.m.	c.p.m. hybridized to SacB	c.p.m. hybridized to SacA
γ -S ATP			
Unbound	7.92×10^6	2,225	430
Bound	2.7×10^5	62	22
γ -S GTP			
Unbound	3.82×10^7	8,732	2,652
Bound	9.67×10^5	108	656

RNA was synthesized in the *in vitro* nuclear transcriptional system described in Table 2 legend with the exception that either γ -S ATP or γ -S GTP was added in place of the respective unsubstituted nucleotide triphosphate. The RNA was extracted with phenol as described and ethanol precipitated three times. The RNA was then dissolved in the buffer (10 mM Tris-HCl, pH 7.9, 100 mM NaCl, 10 mM Na₂ EDTA, 0.1% SDS) and applied to a low cross-linked mercury-agarose column as described elsewhere³¹. The bound material was eluted with TNE buffer containing 10 mM dithiothreitol³⁰. Eluted fractions (0.5 ml) were monitored for radioactivity by removing aliquots (3 µl for γ -S GTP RNA and 12 µl for γ -S ATP RNA) and measuring Cerenkov radiation in 1.5-ml Eppendorf tubes. Peak fractions were pooled, made 50 µg ml⁻¹ with tRNA and precipitated overnight with two volumes of ethanol at 20 °C. The RNA synthesized with γ -S GTP was the result of four identical *in vitro* reactions pooled before the extraction of RNA. Hybridizations were performed to filters stacked in scintillation vials as described in Table 2 legend. Experiments were performed with triplicate samples for γ -S GTP. The largest variation between homologous filters in each hybridization was less than 20%. Backgrounds were reproducibly 10 c.p.m. per 10⁶ c.p.m. input and have been subtracted from the values presented above.

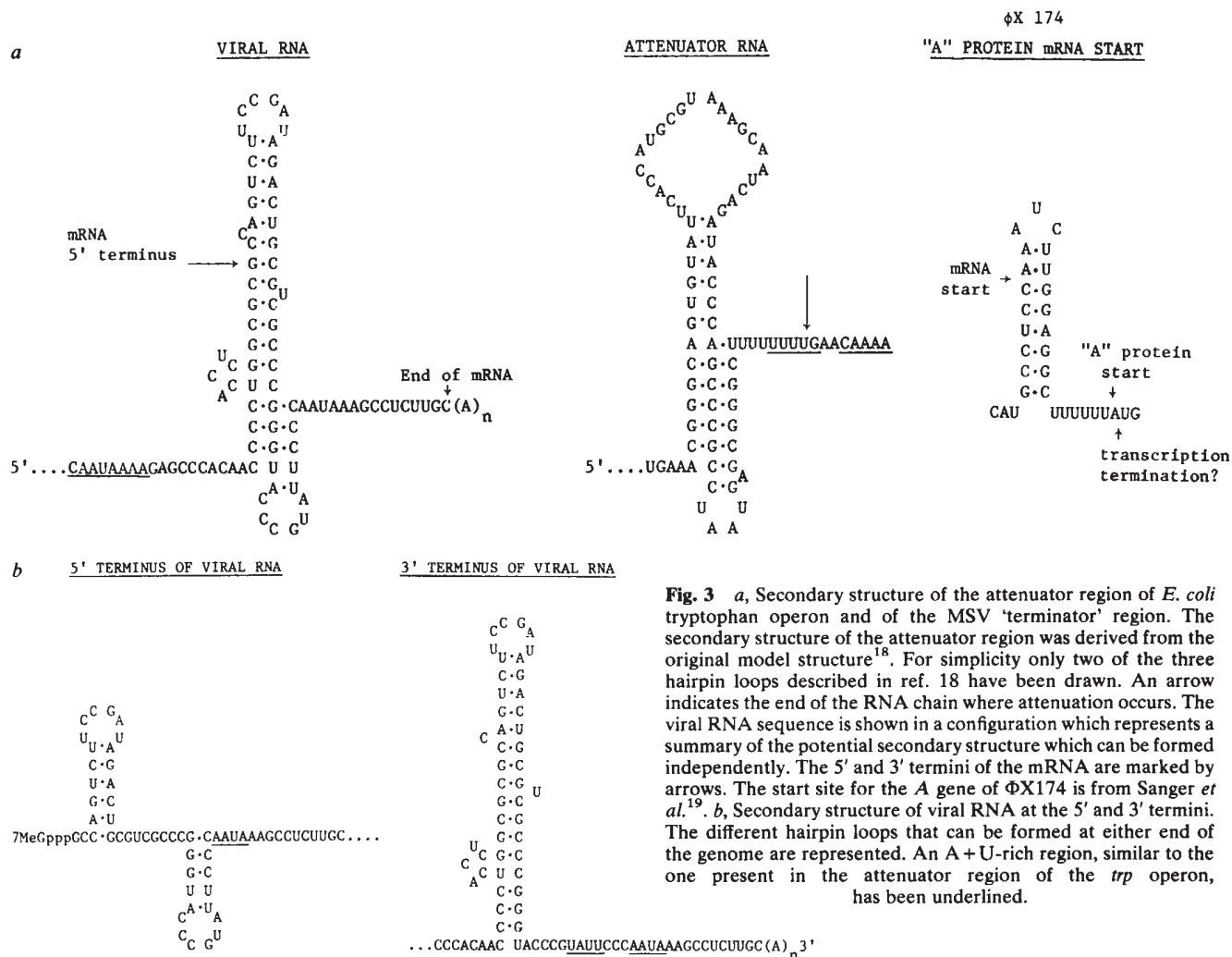


Fig. 3 *a*, Secondary structure of the attenuator region of *E. coli* tryptophan operon and of the MSV 'terminator' region. The secondary structure of the attenuator region was derived from the original model structure¹⁸. For simplicity only two of the three hairpin loops described in ref. 18 have been drawn. An arrow indicates the end of the RNA chain where attenuation occurs. The viral RNA sequence is shown in a configuration which represents a summary of the potential secondary structure which can be formed independently. The 5' and 3' termini of the mRNA are marked by arrows. The start site for the A gene of φX174 is from Sanger *et al.*¹⁹. *b*, Secondary structure of viral RNA at the 5' and 3' termini. The different hairpin loops that can be formed at either end of the genome are represented. An A + U-rich region, similar to the one present in the attenuator region of the *trp* operon, has been underlined.

Secondary structure formed by nascent RNA chains has been implicated in the modulation of termination in bacterial operons^{18,33,34} and phage transcriptional units¹⁹ and might also be involved in MSV. Figure 3*a* shows the proposed structure of the MSV mRNA start site, the *Escherichia coli* *trp* operon attenuator region and the gene A mRNA start region in φX174. It has been postulated that the availability of a specific RNA sequence for loop formation in the *trp* operon and the gene A mRNA start region determines whether or not termination will occur^{18,19}. A similar mechanism may be involved in MSV transcriptional readthrough. In fact, a striking difference between the 5' terminus of virion RNA and its 3' terminus is that sequences between the Hogness box and the capping site are present only at the 3' end of the viral RNA, thus changing the potential for forming secondary structures in the RNA chains. Figure 3*a* and *b* show the potential secondary structures based on the nucleotide sequence. The double-loop structure for the viral RNA represents a summary of the thermodynamically stable hairpin loops in the area between the Hogness box and the 3' terminus of the viral RNA. The left-side loop has an estimated stability of -22 kcal and the smaller right-side loop has a stability of -5 kcal (ref. 35). Both viral and bacterial hairpin loops are formed by a long series of G·C base pairs which would confer a high degree of stability to the loop potentially involved in termination. The potential secondary structure at the 5' terminus and 3' terminus are displayed in Figure 3*b*.

Conclusions

We have presented data which indicate that *in vitro*, the majority of virus-specific transcripts begin at a site which is on a proviral DNA as opposed to a promoter site on cellular DNA. Sequences

which may serve a promoter function—a Hogness box—were found on viral DNA, 26 base pairs upstream from the site corresponding to the first nucleotide of the viral RNA transcript. As the 3' terminus of the mature viral RNA corresponds to a position only 59 nucleotides downstream from the start site, information which conveys the signals for initiation and termination of viral transcription are in close proximity and present on either end of the proviral DNA. These data suggest the necessity of a readthrough mechanism.

Previously, it has been argued that the only set of sequences common to most poly(A)-terminated transcripts was AAU-AAA²⁴. Although the sequence may be involved in polyadenylation, it cannot by itself code for RNA chain termination because it is absent in histone transcripts^{36,37} and is not recognized in the middle of coding sequences of SV40 and BK viruses^{38,40}. Unlike MSV, in Rous sarcoma virus, a chicken retrovirus, the AAUAAA is not contained within the short terminal redundancy and is therefore present only at the 3' end of the transcript (D. Schwartz, R. Tizard and W. Gilbert, personal communication).

Potential secondary structure may differentiate the 3' and 5' ends of viral RNA transcripts and may be involved in regulating transcription. On the other hand, we cannot completely rule out the possibility that transcription termination on MSV proviral DNA is an unregulated process—that 'correct' termination occurs only on a fraction of RNA chains. The result of this unregulated transcription would be the production of at least four specific nuclear RNAs: (1) a 59-base long product, which would be the result of premature termination; (2) mature viral RNA; (3) readthrough products larger than viral RNA; (4) transcripts beginning at the right-end viral promoter and

containing cellular sequences. Note also that our model is not incompatible with the possibility that mature viral RNA is processed from a larger precursor. In this case the large stem and loop structure at the 3' terminus would serve as a recognition site for processing. This model can be tested experimentally. Transfection of cells with plasmids containing the LTR DNA, singly and in tandem repeats, as well as subclones containing specific deletions, should offer a useful approach for studying these phenomena.

We thank J. Pyati for technical assistance, J. Hurwitz for his support and advice, and A. Maxam, D. Schwartz and J. Sims for

advice and for communicating results before publication. This work was supported by NIH grants CA 24223, CA 26860, CA 13330 and GM 13344. D.D. is a recipient of a Junior Research Faculty Award from ACS; B.N.G. is an Irma T. Hirsch scholar and recipient of a Research Faculty Award from ACS; R.M.W. is a recipient of a NIH postdoctoral fellowship.

Since this manuscript was submitted, the nucleotide sequences of the LTRs of several retroviruses have been reported: by Dhar *et al.*⁴¹, Sutcliffe *et al.*⁴², Czernilofsky *et al.*⁴³ and Shimotohno *et al.*⁴⁴.

Received 28 April; accepted 29 October 1980.

1. Bishop, J. M. & Varmus, H. E. in *Cancer, A Comprehensive Treatise* Vol. 2 (ed. Becker, F. F.) (Plenum, New York, 1975).
2. Monroy, G., Jacquet, M., Groner, V. & Hurwitz, J. *Cold Spring Harb. Symp. quant. Biol.* **39**, 1033–1041 (1974).
3. Jacquet, M., Groner, V. & Hurwitz, J. *Proc. natn. Acad. Sci. U.S.A.* **71**, 2045–2049 (1974).
4. Rymo, L., Parsons, J. T., Coffin, J. M. & Weissmann, C. *Proc. natn. Acad. Sci. U.S.A.* **71**, 2782–2786 (1974).
5. Dinowitz, M. *Virology* **66**, 1–9 (1975).
6. Bishop, J. M. *et al.* in *Animal Virology* (eds Baltimore, D., Huang, A. & Fox, C. F.) 1–20 (Academic, New York, 1976).
7. Brugge, J. S., Purchio, A. F. & Erikson, R. L. *Virology* **83**, 27–33 (1977).
8. Fan, H. *Cell* **11**, 297–305 (1977).
9. Hsu, T. W., Sabran, J. L., Mark, G. E., Guntaka, R. V. & Taylor, J. M. *J. Virol.* **28**, 810–818 (1978).
10. Hughes, S. H. *et al.* *Cell* **15**, 1397–1410 (1978).
11. Benz, E. W. Jr & Dina, D. *Proc. natn. Acad. Sci. U.S.A.* **76**, 3294–3298 (1979).
12. Sabram, J. L. *et al.* *J. Virol.* **29**, 170–178 (1979).
13. Gilboa, E. *et al.* *Cell* **16**, 863–874 (1979).
14. Yoshimura, F. & Weinberg, R. A. *Cell* **16**, 323–332 (1979).
15. Shank, P. R. *et al.* *Cell* **15**, 1383–1395 (1978).
16. Haseltine, W. A., Coffin, J. M. & Hageman, T. G. *J. Virol.* **30**, 375–383 (1979).
17. Dina, D. & Benz, E. W. Jr *J. Virol.* **33**, 377–389 (1980).
18. Oxender, D. L., Zurawsky, G. & Yanofsky, C. *Proc. natn. Acad. Sci. U.S.A.* **76**, 5524–5528 (1979).
19. Sanger, F. *et al.* *Nature* **265**, 687 (1977).
20. Gannon, F. *et al.* *Nature* **278**, 428–434 (1979).
21. Konkel, D., Tilghman, S. & Leder, P. *Cell* **15**, 1125–1132 (1978).
22. Baker, C. C., Herisse, J., Courtois, G., Galibert, F. & Ziff, E. *Cell* **18**, 569–580 (1979).

23. Grosschedl, R. & Birnstiel, M. L. *Proc. natn. Acad. Sci. U.S.A.* **77**, 1432–1436 (1980).
24. Benoist, C., O'Hare, K., Breathnach, R. & Chambon, P. *Nucleic Acids Res.* **8**, 127–142 (1980).
25. Lovinger, G. G. & Schoechezman, G. *Cell* (in the press).
26. Lovinger, G. G. & Schoechezman, G. *J. Virol.* **32**, 803–811 (1979).
27. Dina, D. & Beemon, K. *J. Virol.* **23**, 524–532 (1977).
28. Wang, L. H. & Duesberg, P. H. *J. Virol.* **14**, 1515–1529 (1974).
29. Smith, M. M., Reeve, A. E. & Huang, R. C. C. *Biochemistry* **17**, 493–500 (1978).
30. Smith, M. M., Reeve, A. E. & Huang, R. C. C. *Cell* **15**, 615–626 (1978).
31. Wydro, R. M., Brot, N. & Weissbach, H. *Archs Biochem. Biophys.* **201**, 73–80 (1980).
32. Coffin, J. M., Hageman, T. C., Maxam, A. M. & Haseltine, W. A. *Cell* **13**, 761–773 (1978).
33. Keller, E. B. & Calvo, J. M. *Proc. natn. Acad. Sci. U.S.A.* **76**, 6186–6190 (1979).
34. Johnston, H., Barnes, W., Chumley, F., Bossi, L. & Roth, J. *Proc. natn. Acad. Sci. U.S.A.* **77**, 508–512 (1980).
35. Tinoco, I. Jr *et al.* *Nature new Biol.* **246**, 40 (1973).
36. Sures, I., Lowry, J. & Kedes, L. H. *Cell* **15**, 1033–1044 (1978).
37. Busslinger, M., Portmann, R., Irminger, I. C. & Birnstiel, M. L. *Nucleic Acids Res.* (in the press).
38. Fiers, W. *et al.* *Nature* **273**, 113–120 (1978).
39. Reddy, V. B. *et al.* *Science* **200**, 494–502 (1978).
40. Seif, I., Ghoury, G. & Dhar, R. *Cell* **18**, 963–978 (1979).
41. Dhar, R., McClements, W. L., Enquist, L. W. & Vande Woude, G. F. *Proc. natn. Acad. Sci. U.S.A.* **77**, 3937–3941 (1980).
42. Sutcliffe, J. G., Shinnick, T. M., Verma, I. M. & Lerner, R. A. *Proc. natn. Acad. Sci. U.S.A.* **77**, 3302–3306 (1980).
43. Czernilofsky, A. P. *et al.* *Nucleic Acids Res.* **8**, 2967–2984 (1980).
44. Shimotohno, K., Mizutani, S. & Temin, H. M. *Nature* **285**, 550–554 (1980).
45. Vande Woude, G. F. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **76**, 1677–1681 (1979).
46. Canaani, E. & Aaronson, S. A. *Proc. natn. Acad. Sci. U.S.A.* **76**, 1677–1681 (1979).
47. Grunstein, M. & Hogness, D. S. *Proc. natn. Acad. Sci. U.S.A.* **72**, 3961–3965 (1975).
48. Maxam, A. M. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **74**, 560–564 (1977).

Crystal structure of yeast tRNA^{Asp}

D. Moras, M. B. Comarmond, J. Fischer, R. Weiss & J. C. Thierry

Laboratoire de Cristallographie (ERA 08), Institut Le Bel, Université Louis Pasteur, 4 rue Blaise Pascal, 67070 Strasbourg Cedex, France

J. P. Ebel & R. Giegé

Laboratoire de Biochimie, Institut de Biologie Moléculaire et Cellulaire du CNRS, 15 rue Descartes, 67084 Strasbourg Cedex, France

Two independent, three-dimensional structures of yeast tRNA^{Asp}, mainly differing by the conformation of the D loop, have been obtained from a multiple isomorphous replacement (MIR) X-ray analysis at 3.5-Å resolution. The folding of the ribose-phosphate backbone is similar to that found for tRNA^{Phe}; major differences concern the relative positioning of the acceptor and anticodon stems, and the conformation of the loops in the two molecules. Crystal packing involves self-complementary GUC anticodon interactions.

TRANSFER RIBONUCLEIC ACIDS are versatile macromolecules which interact with many components of the protein-synthesizing machinery. Some of the functions of these molecules, like the interaction with the ribosomes, are common to all tRNAs^{1,2}; others, like the specific attachment to the aminoacyl-tRNA synthetases, are specific to each particular tRNA¹. It is therefore expected that all these molecules possess structural similarities at the three-dimensional level, but also differences responsible for their specific functions. To understand the molecular basis of the multifunctional character of tRNA, knowledge of the structure of different tRNAs is necessary.

The first molecular structure of a tRNA obtained from electron-density maps at atomic resolution^{3–6}, that of yeast tRNA^{Phe}, made possible the elucidation of the folding of the polynucleotide backbone which accommodates the clover-leaf domains in a stable L-shaped structure, the structural justification for most of the invariant residues, and the evidence of tertiary interactions necessary to the stability of the structure.

These features, which might be found in all tRNAs, probably account for the general tRNA functions. The fact that similar information was obtained from two independent structures of tRNA^{Phe}, in two different crystal forms, ruled out the possibility of distortions induced by packing forces and gave more confidence for the generality of the L-shaped structure. However, the structural data obtained on yeast tRNA^{Phe} did not allow definitive conclusions about the features responsible for the specific functions of the tRNAs. Furthermore, the nature of conformational changes in tRNA, observed by various methods, is still uncertain^{7,8}. These facts prompted study of the crystallization of tRNAs presenting structural and functional differences compared with tRNA^{Phe}. Emphasis was given to obtaining good crystals of the initiator tRNAs and of tRNAs with short or long extra-loops, or of tRNAs possessing three base pairs in the D stem. In 1977, we succeeded in growing crystals of yeast tRNA^{Asp}, which is a tRNA with a short extra-loop, diffracting better than 3.5 Å of resolution⁹.

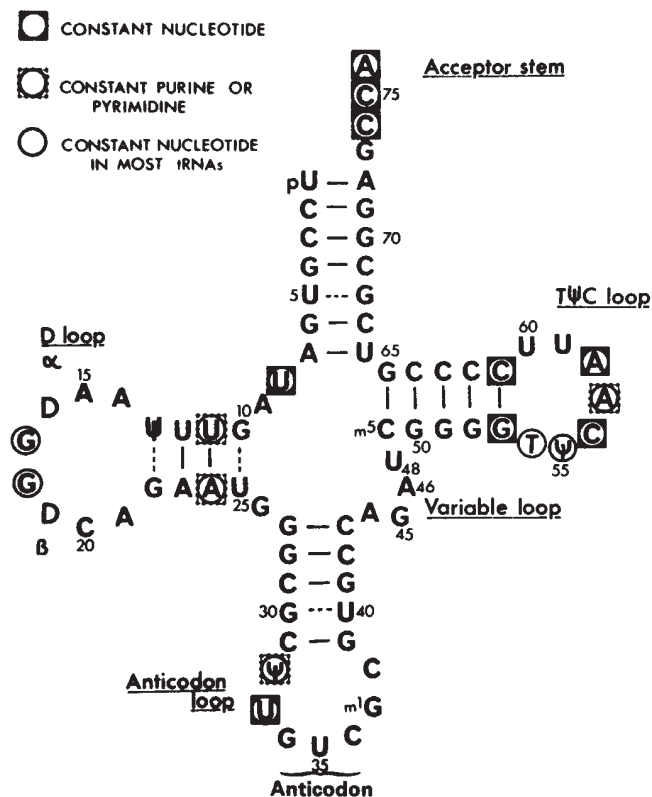


Fig. 1 The nucleotide sequence of yeast tRNA^{Asp} (ref. 13). For convenience the numbering system of the nucleotides is that of yeast tRNA^{Phe}; in the 75-nucleotide long tRNA^{Asp}, position 47 in the variable loop has been omitted. Non-classical Watson-Crick base pairs are indicated by broken lines. The invariant and semi-invariant positions (except for initiator tRNAs) and the constant nucleotides in most tRNAs are indicated. Sequence data are taken from the review of Dirheimer *et al.*²⁶. The α and β regions correspond to the parts of the D loop located respectively at the 5' and 3' side of the constant GG sequence.

It was not until 1979 that new structural information on different tRNAs appeared. In yeast tRNA^{Met}, which was crystallized in high-salt conditions, only a difference in the position of the anticodon stem with respect to tRNA^{Phe} was apparent, but the resolution of the map does not allow a firm conclusion¹⁰. For the structure of *Escherichia coli* tRNA^{Met}, solved using the molecular replacement method, changes from the tRNA^{Phe} model were observed which may be of functional importance¹¹. Yeast tRNA^{Gly} is an exception with an unfolded structure¹², but this tRNA was crystallized at high temperature in the presence of high amounts of organic solvent.

We report here the first structural results on another elongator tRNA, yeast tRNA^{Asp}, based on two independent maps obtained by the MIR method from two interconvertible crystal forms diffracting at high resolution.

Primary structure

The nucleotide sequence of yeast tRNA^{Asp} (ref. 13, Fig. 1) presents some distinctive features compared with yeast tRNA^{Phe} and other tRNAs studied by crystallographic methods. An exceedingly high number (nine out of 10) of GC base pairs in the TΨC and anticodon stems confers an important stability to these regions, as confirmed by solution studies^{14,15}. Conversely, the D stem does not contain any GC base pair and is therefore particularly unstable. Due to this, the number of base pairs in the D stem of tRNA^{Asp} is not well defined and it has long been considered as belonging to class II according to Levitt's tRNA classification¹⁶. Yeast tRNA^{Asp} is also one of the tRNAs containing the fewest number of modified nucleotides: besides the classical D, Ψ and T residues, it only possesses one m¹G and

one m⁵C. Concerning the general organization of the tRNA^{Asp} clover-leaf, the α and β regions in the D loop (three bases each) differ from those in tRNA^{Phe} (four and two bases for the α and β regions respectively) and the length of the variable loop (four bases) is shorter than in tRNA^{Phe} (five bases). Another difference between the two tRNAs is the replacement at position 73 of A in tRNA^{Phe} by G in tRNA^{Asp}. This position has been proposed as a discriminator site for tRNA aminoacylation¹⁷. Finally, tRNA^{Asp} possesses a self-complementary GUC anticodon¹⁸ which renders it a tempting model system for studying tRNA-mRNA interactions.

Crystal analysis

We used 150 mg of pure tRNA^{Asp}, purified from brewer's yeast tRNAs by countercurrent distribution followed by benzoylated DEAE-cellulose chromatography. For crystallization, tRNA^{Asp} was used as a sodium salt, prepared by dialysis against 40 mM sodium cacodylate buffer containing 2 mM MgCl₂. Crystals were obtained in 62% ammonium sulphate using vapour diffusion techniques⁹. The temperature was a decisive parameter in crystal growth. Crystals suitable for diffractometer measurements were obtained, although in an exceedingly poor yield, by careful monitoring of the concentration of ammonium sulphate against temperature. Before being exposed to X rays, all crystals were stabilized against 70% ammonium sulphate solutions. The space group is C222₁, with one molecule of tRNA in the asymmetric unit. The solvent content of the crystals, ~75%, is comparable with that of the orthorhombic crystals of yeast tRNA^{Phe}. At 20 °C, the temperature at which the best crystals are grown, two crystal forms, nonisomorphous and interconvertible, are obtained.

$$\begin{array}{ll} a = 61.5 \text{ \AA} & a = 60.3 \text{ \AA} \\ \text{(A) } b = 67.5 \text{ \AA} & \rightleftharpoons \text{(B) } b = 68.0 \text{ \AA} \\ c = 149.5 \text{ \AA} & c = 149.5 \text{ \AA} \end{array}$$

The mean difference in the native F between the two crystal forms ($\Sigma|\Delta F|/\Sigma|F|$) is of 15% throughout the $\sin \theta/\lambda$ range. As it was not possible to produce selectively and identify one form or the other, two independent structure determinations had to be carried out by the MIR method. The search for the heavy-atom derivatives was hampered by the interconvertibility of the two crystal forms. For example, it was difficult to obtain a gold derivative of form A, as addition of gold salt drives crystals to an intermediate form close to that of crystals B. Experimental details are given in Fig. 2 legend.

Electron density map

The shape and limits of the tRNA^{Asp} molecule in the crystallographic cell could be drawn from an earlier 5-Å resolution electron-density map phased with two heavy-atom derivatives (Gd and Au, f.o.m. = 0.72). This map revealed an L-shaped structure roughly comparable with that found for tRNA^{Phe}. Some characteristic features of the tRNA structure, like the helical stems and the anticodon loop, were clearly recognizable. The region situated near the D and TΨC stems lies close to a 2-fold axis parallel to the crystallographic direction a , which brings two molecules in close contact in that part of the cell. In this case, the problem of the boundary could be solved by topological considerations and continuity of the helices. At that resolution, however, the precise frontier between the TΨC loop and the CCA end of another molecule could not be unambiguously resolved due to the packing conditions (see below).

The 3.5-Å resolution data confirmed this first analysis and enabled us to trace the complete ribose-phosphate backbone. A few sections of the resulting electron-density map are presented in Fig. 3. Figure 3a shows a portion of the anticodon loop viewed along the direction of the a axis. Figure 3b represents part of the TΨC stem viewed down the helical stem which runs parallel with the a axis. In both, the locations of the phosphate groups are indicated by dots and correspond to bulbs of quasi-spherical electron density. A striking feature of this map is the high

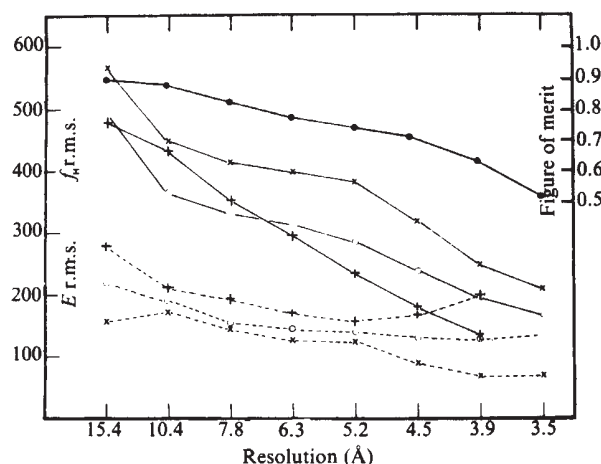


Fig. 2 Refinement statistics of form A. The diffracted intensities were collected on a four-circle Philips PW1100 diffractometer modified for protein work. Great care was taken concerning the decay of intensities in X-ray exposure, this effect being related to the Bragg angle. In standard conditions, a native crystal loses 1% of intensity per h above 3.5-Å resolution. For each crystal, data were collected in successive shells of decreasing resolution, the starting value being determined by the angle to which significant diffraction was observed. For example, for the native data of form B, six crystals were used, scaled together and merged; 5,262 independent significant reflections to 3-Å resolution were then kept, out of 17,440 observed intensities ($F > 2\sigma F$), the agreement factor of intensities being equal to 5.5%. Details concerning the measurement and treatment of the data will be published elsewhere. The best heavy-atom derivative was obtained with gadolinium. The observed unicity of the binding site, when varying the Gd concentration, can be partly explained by the crystallization conditions. Lutetium and lanthane lead to similar results. Gold has two binding sites. The root mean square (r.m.s.) calculated heavy-atom structure amplitudes (f_H , continuous line) and the r.m.s. lack of closure (E , broken line) are plotted against resolution for the different isomorphous derivatives used in the refinement: $\times$, $Gd_2(SO_4)_3$; $+$, $Au(en)_2Cl_3$; and $\circ$, the mixed derivative. The distribution of the figure of merit is also indicated ($\bullet$), the overall value being 0.62 for 3,477 phased reflections.

contrast between the electron density of the molecule and the solvent regions. This is even true, but to a lesser extent, for most of the loop regions which are fortunately stabilized by appropriate intermolecular contacts. Except for part of the D loop, the phosphate groups can be identified as the peaks of highest density roughly 6-Å apart. The high content of ammonium sulphate in the crystals with the resulting high ionic strength does not affect the observed signal-to-noise ratio. The background barely reaches the first contour level except in the groove formed by the T ψ C loop where a higher density can be observed which is not yet related to any molecular feature. The

electron density is also perturbed near the heavy-atom sites where ripples of negative density exist.

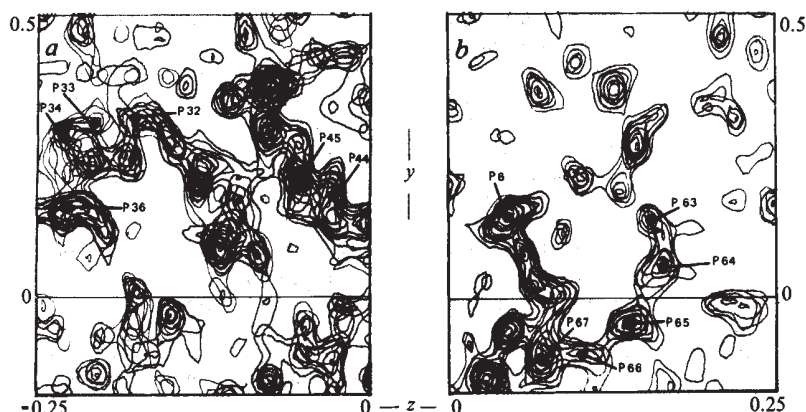
The very high proportion of GC base pairs in tRNA^{Asp} probably has an important effect on the stability of the double-helical stems. This in turn is reflected in the experimental electron density. In most places the phosphate groups can be distinguished from the riboses, which appear as flat, connecting densities between the phosphate bulbs. The electron density corresponding to the paired bases can clearly be seen in the double-helical regions; this is also true for many bases of the loop regions which are probably involved in tertiary interactions, either intra- or intermolecular. This is obviously the case for the electron density associated with the T ψ C and anticodon loop regions, especially the latter which is stabilized by intermolecular contacts across a 2-fold axis. The interactions with the GUC anticodon of another molecule restrain the motion of the atoms. The resulting electron density is thus remarkably strong. At the other extremity of the map, the single-stranded CCA end is clearly visible, due to stabilizing interactions with the T ψ C loop of the neighbouring molecule. The situation is different in the region of the map corresponding to the D loop which seems weaker in both A and B forms. This was also noticed by Robertus *et al.* in the 3-Å resolution map of the monoclinic form of yeast tRNA^{Phe} (ref. 5).

Model

A stereo picture of a balsawood model built out of the 3.5-Å resolution electron-density map of form A is shown in Fig. 4a. This picture corresponds to a view down the crystallographic *b* axis. The two major structural units giving the molecule its L shape can be recognized: vertically, from the bottom, the anticodon stem followed by the D stem; horizontally, from the corner of the molecule, the T ψ C with the amino acid-accepting stem stacked on it. The molecule is approximately 20-Å thick in the viewer's direction. This model can be compared with both balsawood and space-filling models of yeast tRNA^{Phe} (refs 19, 20). It clearly shows the more open conformation of the L in tRNA^{Asp}, which resembles a boomerang. Two major grooves, sharply cut and without any significant electron density, can be seen in the two limbs of the molecule. These are characteristic of a RNA-type double helix.

The entire ribose-phosphate backbone of tRNA^{Asp} deduced from the 3.5-Å resolution maps is presented in Fig. 4b in the same perspective as the balsawood model (ORTEP drawing²¹). The crude position of the peaks of highest density, approximately 6-Å apart from each other, was used as coordinates for the phosphate groups. The overall folding agrees with that observed for tRNA^{Phe}. This extends the validity of the tRNA^{Phe} model to the short-loop tRNA family and provides direct evidence that the overall folding of a tRNA remains unchanged in high (tRNA^{Asp}) or low (tRNA^{Phe}) salt concentrations. Even the D stem is clearly visible despite the weakness of its base pairing (two AU and two GU base pairs). This apparently contradicts

Fig. 3 Two composites of the 3.5-Å MIR map of yeast tRNA^{Asp}, form A. *a* Shows the electron density corresponding to part of the anticodon loop and stem (four sections between $x = -7.7$ and -3.7 Å). This region is characterized by the absence of any significant density in the major groove of the anticodon stem, and by the high electron density of the anticodon loop when compared with that of the nearby helical stem. *b* Shows densities in the amino acid acceptor- and T ψ C-stem regions (two sections between $x = 19.4$ and 20.7 Å). The electron density corresponding to a Watson-Crick base pair (G6 C67) is seen between P6 and P67.



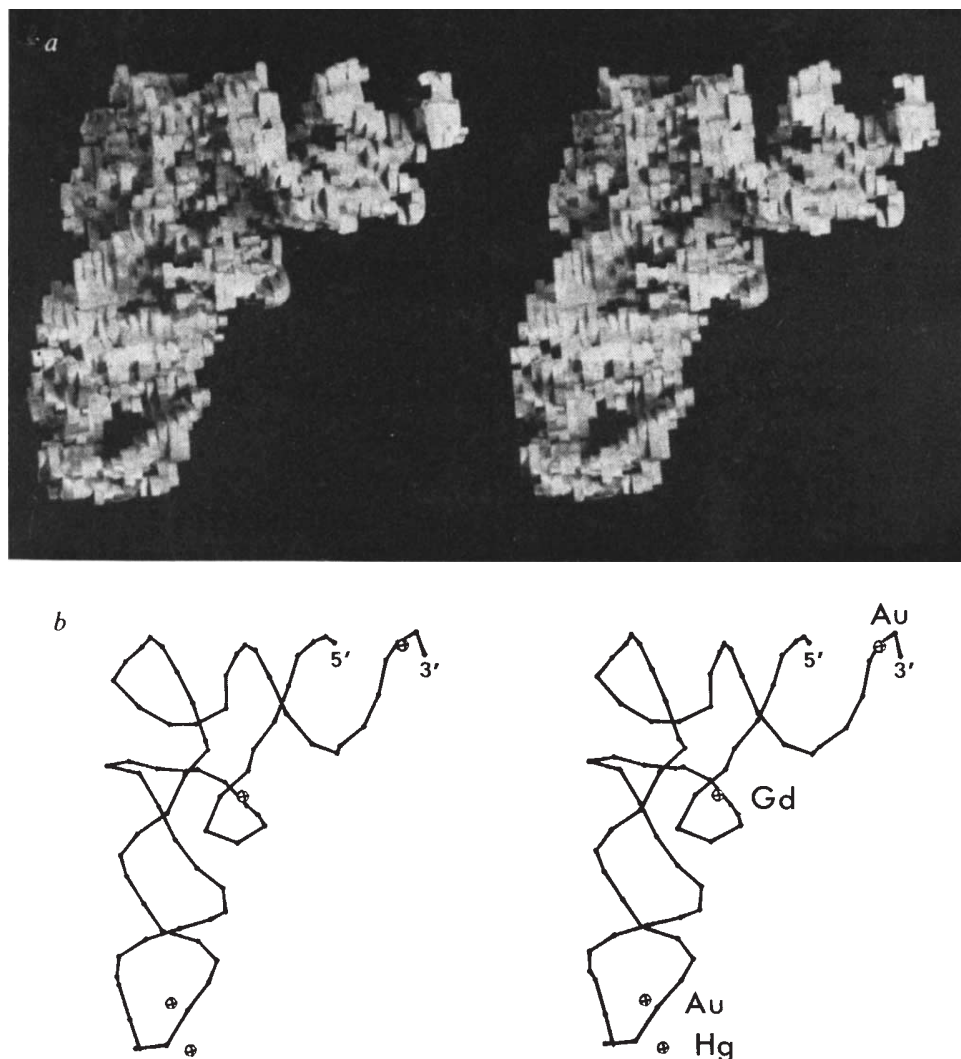


Fig. 4 *a*, Stereo picture of a balsawood model of tRNA^{Asp} built out of the 3.5-Å resolution electron-density map of form A. The view is down a direction roughly parallel to the crystallographic *b* axis. In that direction the thickness of the molecule is approximately 20 Å. *b*, Stereo view (ORTEP drawing) of the ribose-phosphate backbone presented in the same perspective as the balsawood model. The positions of the heavy-atom binding sites are indicated.

former NMR studies where these interactions, essentially the two AU base pairs, could not be seen in normal ionic conditions; but this is most likely related to the particular distribution of GC and AU base pairs in tRNA^{Asp} (ref. 15).

The positions of the heavy-atom binding sites on tRNA^{Asp} are also shown in Fig. 4*b*. The lanthanide site is situated near the position of phosphates U8 and A9. It is similar to one of the two strong samarium binding sites in yeast tRNA^{Phe} (ref. 22). The distance between the gadolinium site and the phosphates 8 and 9 is ~4 Å; although this is not a refined value, it strongly suggests a direct interaction between the lanthanide and these phosphate groups, as observed in tRNA^{Phe} . The fact that this is the unique marked site in tRNA^{Asp} , whereas three other binding positions are observed in tRNA^{Phe} , can be related to the high structural selectivity of the complexing cavity for lanthanides. In the crystallization medium of tRNA^{Asp} , these cations have two strong competitors: magnesium and ammonium. Two possible explanations for the different behaviour of lanthanide cations observed in tRNA^{Asp} could be the strong competition of ammonium for weak magnesium binding sites and/or a different D-loop conformation in tRNA^{Asp} for the second, strong magnesium binding site found in tRNA^{Phe} . Gold binds in two different sites with different affinities. The major site lies near the CCA end (phosphates C75 and A76), thus being indicative of an interaction with the phosphate group, but binding to a base of the TΨC loop of a neighbouring molecule cannot be excluded as the two parts of the molecule are close. The second gold binding site is in the pocket formed by the anticodon loop and most likely involves an attachment to the bases. Note that none

of these gold binding sites is observed in yeast tRNA^{Phe} (ref. 23). For a mercury derivative, not used in the phasing, the heavy atom is situated on the 2-fold axis relating the anticodons of two equivalent molecules. It is a good candidate for binding to U35.

Structural comparison with yeast tRNA^{Phe}

Earlier attempts to solve the crystallographic structure of yeast tRNA^{Asp} by the molecular-replacement method failed using the structure of yeast tRNA^{Phe} as a model. Although the major peak of the rotation function was close to the best possible compromise, no reasonable solution could be obtained from the various translation functions tested. It was only after obtaining an independent chain tracing of tRNA^{Asp} from the MIR map and having realized the close similarity of folding with that of tRNA^{Phe} that it was possible to position roughly the tRNA^{Phe} model in the tRNA^{Asp} cell. A 6-Å resolution set of observed structure factors (F_o) was then used to obtain the best fit of the two molecules by comparison with the calculated values (F_c) derived from the tRNA^{Phe} model. An optimization of the six parameters necessary to position the tRNA^{Phe} model (the three eulerian rotation angles and the coordinates of the centre of mass of the molecular electron density) was carried out by minimization of $(F_o - F_c)^2$ quantities by a least-square procedure. Convergence was slow; the reliability factor $R = \sum ||F_o| - |F_c|| / \sum |F_o|$ decreased from 0.55 to 0.50 in 19 cycles. The high final *R* value for 6-Å resolution data is indicative of poor agreement between the model and the observed data. Details on these procedures will be published elsewhere.

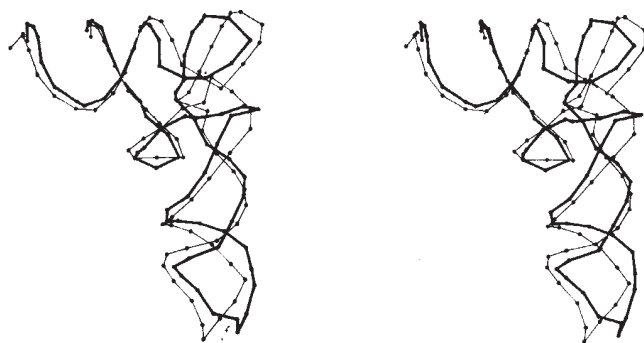


Fig. 5 Comparison of the ribose-phosphate backbones of yeast tRNA^{Asp} in form A (heavy lines) with that of yeast tRNA^{Phe} (light lines).

The resulting parameters were used to trace a superposition of the two tRNA backbones shown in Fig. 5. We recall that the tRNA^{Asp} is an independent, unrefined model resulting from a crude positioning of phosphate groups. The first overwhelming impression concerns the similarity of the overall folding of both tRNAs. Careful inspection, however, clearly uncovers differences, but at the present level of accuracy only the major ones can be underlined. Figure 5 confirms the more open L conformation of tRNA^{Asp}, although the superposition method tends to minimize the differences. The angle formed by the two limbs of the L is increased by more than 10° (Fig. 6). This results in different positioning of the anticodon and of the TΨC stems and loops with respect to fixed acceptor and D stems which superpose well to the corresponding parts of tRNA^{Phe}; as a result, the D and anticodon stems are more coaxial. It is tempting to propose that conformational changes in tRNAs could result from concerted movements of the clover-leaf domains different from that suggested by the L-shaped model (see Fig. 6). A concerted behaviour of the D and acceptor stems was also suggested by thermodynamic studies¹⁴ which showed that the simultaneous melting of the tertiary structure and of these two regions is the first step of denaturation of tRNA^{Asp}.

A tilt of the anticodon stem is also observed in the initiator tRNA from yeast, but in this tRNA the angle between the acceptor and anticodon stems decreases by ~6° when compared with tRNA^{Phe} (J. Sussmann, personal communication). A possible regulatory effect of spermine on the tilt of the anticodon stem was suggested by Quigley *et al.*²⁴ after discovery of a spermine molecule bound at the hinge region. High salt conditions might have a competitive effect which would induce a structural change at that level. However, the fact that those ionic conditions with similar spermine concentration lead to opposite changes in two different tRNAs greatly weakens such an explanation. In the variable region the deletion of one nucleotide in yeast tRNA^{Asp} does not cause an important change. From the superposition diagram, the missing nucleotide is probably U47, which in tRNA^{Phe} points towards the exterior of the molecule, the rest of the chain just being stretched toward C49. The resulting shift might, of course, trigger the observed tilt of the anticodon stem, thus making it a characteristic of the tRNAs with four bases in the variable region.

The relative position of the TΨC and D loops is different from that observed in tRNA^{Phe}. The precise interactions in this part of the map are not known and further refinements are required. Nevertheless, the interaction scheme between the loops present in tRNA^{Phe} is clearly not valid here. For example, the fact that in tRNA^{Asp} the two conserved guanines G17 and G18 are moved in the sequence with respect to tRNA^{Phe} makes base pairing between G18 and C56 impossible, unless the loop is shifted.

Differences exist even between the two structures of tRNA^{Asp} corresponding to the two crystal forms by comparison of the two maps. An averaged map of the density of the two crystal forms shows a much higher contrast everywhere, except in the D-loop

region which is smeared out between nucleotides 16 and 19. This part of the molecule can be otherwise followed on each individual map. The transition between the two forms occurring around 20°C agrees with the observed difference and stresses the mobility of the D loop. The previously mentioned thermodynamic studies imply that the observed movements of the D loop are not related to a set of especially weak interactions which loosen the structure in that region, as the acceptor stem and the D-stem helix melt simultaneously with a collapse of the tertiary structure.

Intermolecular contacts

Three major intermolecular contacts are involved in the stabilization of the crystal packing of tRNA^{Asp}. One contact already mentioned involves interactions between the anticodons of two molecules related by a 2-fold symmetry axis parallel to the crystallographic *b* direction. It assures the crystalline cohesion in the direction of the crystallographic *c* axis. This contact is particularly important because of the self-complementarity of the anticodon GUC of tRNA^{Asp}. Grosjean *et al.*¹⁸ have already shown that this association exists in solution. This is reflected, as shown in Fig. 3b, by a high value of the corresponding electron density which compares with that of helical regions. The comparison in that region of the tRNA^{Asp} backbone with that of tRNA^{Phe} in the orthorhombic form, where one anticodon is stacked on top of the other, is shown in Fig. 7. Similar stacking interactions in tRNA^{Asp} can be excluded by the distances between the related phosphate groups which suggest a base pairing, although the precise nature of these interactions needs a

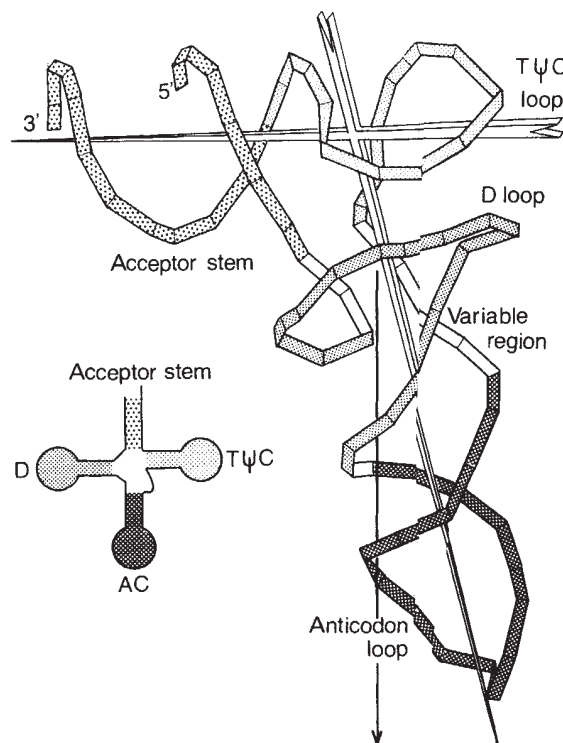


Fig. 6 Diagrammatic representation of the ribose-phosphate backbone and of the cloverleaf domains of tRNA^{Asp} (indicated by different shadings). The two pointers show the more open conformation of tRNA^{Asp} with respect to tRNA^{Phe}. Assuming an exact superposition of the acceptor stem in the two molecules, the shifts which give the tRNA^{Asp} molecule its boomerang shape result mainly from a different positioning of the A (and TΨC) parts of the cloverleaf, the acceptor and D stems being positioned as in tRNA^{Phe}. The arrow represents the direction of the helical axis of the anticodon stem in tRNA^{Phe}.

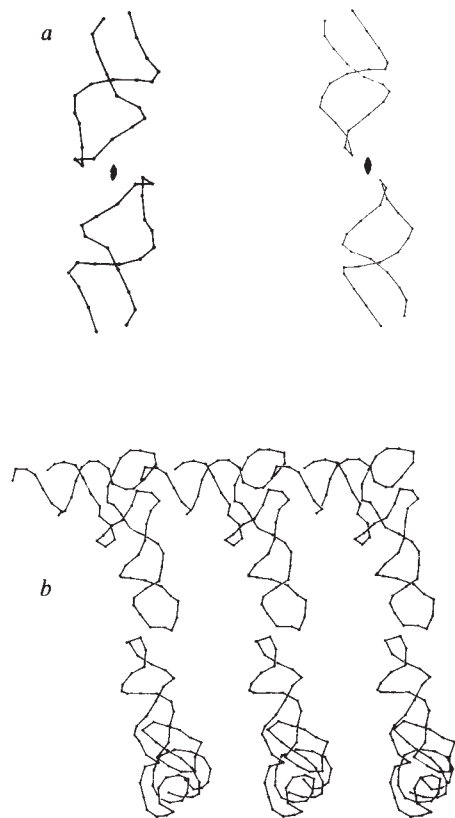


Fig. 7 Packing diagrams. *a*, Anticodon-anticodon interactions. The 2-fold axis-related backbones of the anticodon stem and loop are shown for yeast tRNA^{Asp} (left) and for the orthorhombic form of yeast tRNA^{Phe} (right). *b*, Partial-packing diagram of yeast tRNA^{Asp}. For clarity only four out of the eight equivalent positions of the space group are traced for one and a half unit cells. The drawing shows the anticodon-anticodon contacts and the interactions between the CCA end and the TΨC loop of adjacent molecules; in the crystals this results in a network of two classes of infinite pseudo-helices, 90° apart. The crystallographic 2-fold axis relating the two anticodons is at 45° from the figure plane.

refined model. The two anticodon loops form the two strands of a double-helical structure which might represent a model of interaction between tRNA and mRNA.

Another strong intermolecular contact, introduced by the C-mode translation, involves the TΨC loop and the acceptor stem of a neighbouring molecule. This results in two infinite, kinked helices related by a dyad and 90° apart. Figure 7*b* represents a partial-packing diagram, which underlines the two types of intermolecular contacts described above. A third and quantitatively important intermolecular contact is a consequence of the crystallographic 2-fold axis parallel to *a*. It brings the anticodon stem and the D stem of one molecule in close contact

with the TΨC stem of another one. In this arrangement a cage is formed by the 5' parts of the D stems of the two symmetrically related molecules. The two symmetric gadolinium atoms, 6-Å apart, enter in this cage during isomorphous substitution.

Biological implications

The three-dimensional structure of tRNA^{Asp} confirms the folding originally found in tRNA^{Phe}. This result fits with an unitary structural scheme and is probably a consequence of the existence of tertiary interactions brought by the invariant and semi-invariant residues. This scheme brings together both elongator and initiator tRNAs and integrates many properties of these molecules (that is, interactions on the ribosome^{1,2}, similar types of interaction in various tRNA-synthetase systems¹, mischarging of numerous tRNAs by the same synthetase²⁵). The tRNA^{Asp} structure also presents differences from the canonical tRNA^{Phe} model. With the present information, only large differences can be stressed with confidence. The most important deviations concern the respective position of the acceptor and anticodon arms of the boomerang-shaped molecule and the conformation of the loop regions. It may be asked whether the different conformation of tRNA^{Asp} is indicative (1) of an archetype of another tRNA class or (2) of another conformation of a standard tRNA structure.

Hypothesis (1) would emphasize the structural role of the clover-leaf organization. In that case, the existing differences between the structure of tRNA^{Phe} and the initiator tRNAs on the one hand, and the structure of tRNA^{Asp} on the other, should result from differences in the variable loop and/or in the β region. A similar relationship between these tRNAs came out of incorrect aminoacylation studies²⁵.

Hypothesis (2) would reflect the conformational flexibility of the tRNA molecules. Would then the conformational changes be induced by the interactions between the self-complementary anticodons, mimicking the ribosomal situation, or is it just a consequence of crystallization conditions, namely high salt concentration? The latter possibility seems to be ruled out by the results obtained with the initiator yeast tRNA^{Met}, where similar ionic conditions lead to different structural variations. Probably the induced conformational changes and sequence-bound differences act together to give the observed structural variations. A more accurate model, as well as a detailed comparison of the two electron-density maps available, will increase our knowledge of tRNAs and help answer some of the open questions.

This work was supported by grants from the CNRS (Action Thématique Programmée: Structure Tridimensionnelle des Protéines) and DGRST. We thank M. Delfau for assistance with the purification and crystallization of tRNA^{Asp}, M. Schlegel for help with countercurrent distribution of crude tRNA, and A. Cassou, S. Wechsler and J. Born for technical help. We also thank M. G. Rossmann for supplying the Purdue crystallographic programs from which most of the programs used in this study are adapted, A. Lifschitz for the least-square superposition, G. Quigley, N. Woo and A. Rich for discussions and programs, and D. Blow, G. Dirheimer, B. Jacrot and A. Klug for encouragement and helpful discussions.

Received 30 June; accepted 7 October 1980.

- Schimmel, P. R., Söll, D. & Abelson, J. N. (eds) *Transfer RNA: Structure, Properties and Recognition* (Cold Spring Harbor Monogr. Ser. 9A, New York, 1979).
- Clark, B. F. C. in *Ribosomes: Structure, Function and Genetics*, 413-444 (University Park Press, Baltimore, 1979).
- Kim, S. H. *et al. Science* **185**, 435-440 (1974).
- Sussmann, J. L., Holbrook, S. R., Warrant, R. W., Church, G. M. & Kim, S. H. *J. molec. Biol.* **123**, 607-630 (1978).
- Robertus, J. D. *et al. Nature* **250**, 546-551 (1974).
- Jack, A., Ladner, J. E. & Klug, A. *J. molec. Biol.* **108**, 619-649 (1976).
- Schwarz, U., Manzel, H. M. & Gassen, H. G. *Biochemistry* **15**, 2484-2490 (1976).
- Ehrlich, R., Lefevre, J. L. & Remy, P. *Eur. J. Biochem.* **103**, 145-153 (1980).
- Giegé, R., Moras, D. & Thierry, J. C. *J. molec. Biol.* **115**, 91-96 (1977).
- Schevitz, R. W. *et al. Nature* **278**, 188-190 (1979).
- Woo, N., Roe, B. & Rich, A. *Nature* **286**, 346-351 (1980).
- Wright, H. T., Manor, P. C., Beurling, K., Karpel, R. L. & Fresco, J. L. in *Transfer RNA: Structure, Properties and Recognition*, 145-160 (Cold Spring Harbor Monogr. Ser. 9A, New York, 1979).

- Gangloff, J., Keith, G., Ebel, J. P. & Dirheimer, G. *Nature new Biol.* **230**, 125-127 (1971).
- Coutts, S. M., Gangloff, J. & Dirheimer, G. *Biochemistry* **13**, 3938-3948 (1974).
- Robillard, C. T. *et al. Biochemistry* **15**, 1883-1888 (1976).
- Levitt, M. *Nature* **224**, 759-763 (1969).
- Crothers, D. M., Seno, T. & Söll, D. G. *Proc. natn. Acad. Sci. U.S.A.* **69**, 3063-3067 (1972).
- Grosjean, H., De Henau, S. & Crothers, D. M. *Proc. natn. Acad. Sci. U.S.A.* **75**, 610-614 (1978).
- Rich, A. in *Ribosomes*, 871-884 (Cold Spring Harbor Monogr. Ser. 4, New York, 1974).
- Holbrook, S. R., Sussmann, J. L., Warrant, R. W. & Kim, S. H. *J. molec. Biol.* **123**, 631-660 (1978).
- Johnson, C. K. *ORTEP Rep. ORNL-3794* (Oak Ridge National Laboratory, Oak Ridge, Tennessee, 1965).
- Jack, A., Ladner, J. E., Rhodes, D., Brown, R. S. & Klug, A. *J. molec. Biol.* **111**, 315-328 (1977).
- Stout, C. D. *et al. Acta Crystallogr.* **B34**, 1529-1544 (1978).
- Quigley, G. J., Teeter, M. M. & Rich, A. *Proc. natn. Acad. Sci. U.S.A.* **75**, 64-68 (1978).
- Ebel, J. P. *Biochimie* **55**, 547-557 (1973).
- Dirheimer, G., Keith, G., Sibling, A. P. & Martin, R. P. in *Transfer RNA: Structure, Properties and Recognition*, 19-41 (Cold Spring Harbor Monogr. Ser. 9A, New York, 1979).

The structure of histone H1 and its location in chromatin

J. Allan*, P. G. Hartman†, C. Crane-Robinson† & F. X. Aviles‡

* Biophysics Laboratories, King's College, London WC2B 5RL, UK

† Biophysics Laboratories, Portsmouth Polytechnic, Portsmouth PO1 2QQ, UK

‡ Department of Biochemistry, Autonomous University, Barcelona, Spain

On the basis of their primary structure, the lysine-rich histones are a unified family of proteins. Each has an amino acid chain which falls into three distinct domains. Only the central domain (~80 residues) is in a folded conformation. It is protected from trypsin digestion in chromatin and corresponds to the segment of highest sequence conservation. Without the flanking domains it is able to close two full turns of DNA in the nucleosome and can thus locate the H1 molecule.

THE lysine-rich histones are a group of chromosomal proteins that show considerable sequence variation when contrasted with the core histones. If calf thymus H1 is taken as the archetypal molecule, then rabbit thymus and chicken erythrocyte H1s have sequences very similar to calf¹, and the trout testis H1 sequence shows only a few variations² (Fig. 1). Avian erythrocyte H5 and sea urchin sperm H1 (Φ 1) have been considered as specialized versions of H1 but have sequences clearly related to that of calf thymus H1³⁻⁵. Recent studies of an H1 subfraction from ox liver, H1^o, indicate it to be intermediate between calf thymus H1 and avian erythrocyte H5 (ref. 6). A partial sequence of an H1 molecule from somatic tissue of sea urchin⁷ is also included in Fig. 1. The sequences of both sea urchin H1 molecules show greater homology with H5s, particularly in the central domain, than with the vertebrate H1s. On the basis of primary structure, it is clear that the lysine-rich histones are a unified family.

Studies of extracted lysine-rich histones in free solution have shown conformational homology within the family. Although fully denatured at pH 3 in the absence of salt, they can be induced to fold by increasing the pH and ionic strength, but the folded domain encompasses only a limited portion of the polypeptide chain⁸. For calf thymus H1 the folding domain has been defined as residues 36–121 (ref. 9). Trypsin digestion of histone H5 similarly defines residues 22–100 of that histone as containing all the secondary and tertiary structure¹⁰. Sea urchin sperm H1 also shows a trypsin-resistant peptide of approximately 80 residues¹¹, representing the sequence 36 to about 116 (P. Sautière and P. Puigdomenech, personal communication). The finding that only a limited portion of the chain is folded and trypsin resistant could, however, be a consequence of partial denaturation on extraction and for this reason we have looked to see whether the same domain is protected from digestion in nuclei.

Trypsin digestion of nuclei

Figure 2 shows a time course of digestion of calf thymus nuclei in isotonic conditions. The H1 subfractions are seen to be gradually degraded to a limiting peptide which runs very close to the peptide (36–121) prepared by trypsin digestion of H1 in free solution⁹. The two peptides have very similar amino acid analyses, and high-resolution proton-resonance spectroscopy shows that the tertiary folding of intact calf thymus H1 and both peptides are the same. We conclude that essentially the same peptide is protected from digestion in nuclei as in free solution.

Digestion experiments with proteases to probe conformation frequently show that protected and folded domains are linked by relatively short bridges of susceptible amino acid residues (hinge regions, for example, *lac* repressor). Does the H1 molecule have

only one, short, protected domain while the remainder is fully accessible to trypsin? Figure 2 shows that at early digestion times many peptides have lengths intermediate between intact H1 and the resistant peptide. The two flanking domains are clearly very accessible in chromatin and therefore probably not in a compact, folded conformation.

Correlation of primary structure and trypsin-defined conformation

Comparison of the observed trypsin-resistant regions of three rather distant members of the H1 family with the sequence homology (Fig. 1) indicates that the protected domain coincides with the strongly conserved and more hydrophobic region of the sequence in all three cases. Although the conservation of this central domain is less than that of any of the core histones^{1,5,12}, it is very much greater than that of the flanking domains. The N-terminal domain shows only low sequence homology throughout the family and varies greatly in length from 40 residues in sea urchin sperm to 13 residues in sea urchin embryos. Homology in the C-terminal domain is less clear-cut as in calf thymus H1, alanine + lysine + proline make up 80% of the residues. As there is considerable replacement of lysine by arginine and of alanine by serine in avian H5, some major restructuring is clearly possible in the C-terminal domain. It is clear, however, that the central, conserved sequences correspond closely with the protected peptides of calf H1, chicken H5 and sea urchin sperm H1, and homologous peptides of ~80 residues are obtained from all three proteins.

Three structural domains of H1 are therefore defined and this view of the molecule is shown in idealized form in Fig. 1. What, then, is the function and spatial arrangement of these three domains of H1 in the structure of chromatin?

Experimental approaches to the location and function of H1

H1 has been implicated in chromatin condensation¹³⁻¹⁷. We have tested this function by preparing oligonucleosomes depleted of lysine-rich histone¹⁸. To this were reassociated peptides of H1 and the sedimentation coefficient of the reconstituted material compared with a control reconstituted with intact H1. Both reassociation and sedimentation were at 80 mM NaCl, conditions in which the extended nucleofilament is almost fully compacted into the higher-order structure¹⁷. Previous results have shown¹⁹ that a ratio, r , of 2 moles of intact H1 per nucleosome is required to restore the S-value of the native chromatin and above this value extensive precipitation is

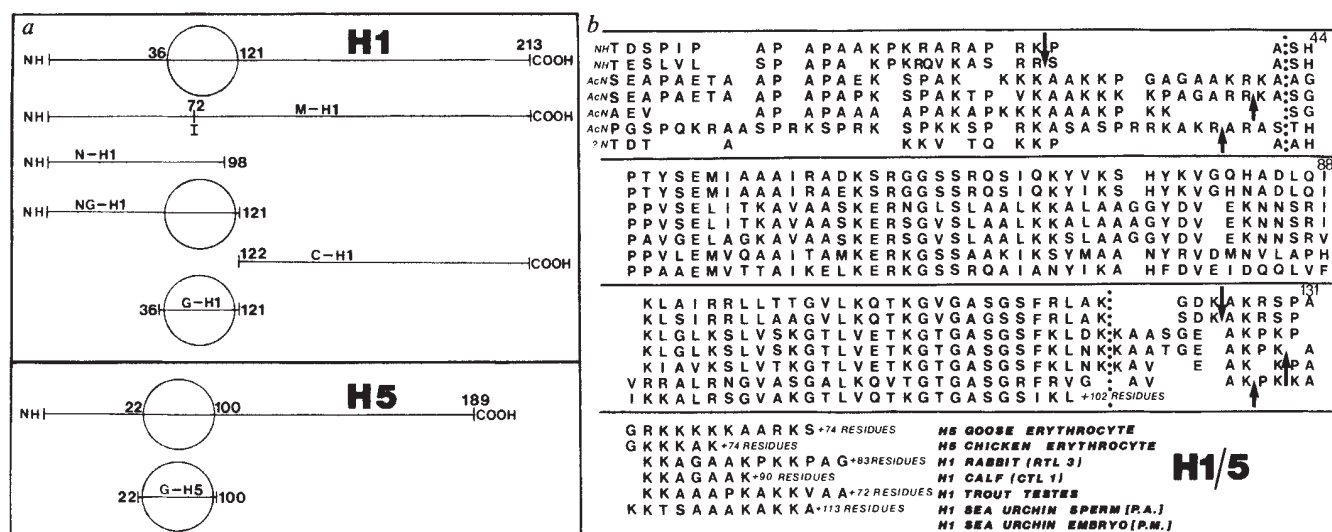


Fig. 1 *a*, Sequences of several lysine-rich histones aligned for maximum homology. The C-terminal part, consisting largely of proline, alanine and lysine (arginine in H5s), is not fully shown. Vertical dotted lines represent the proposed limits of the strongly conserved globular domain. Bold arrows indicate tryptic cutting points in free-solution digestion of histones at high ionic strength (refs 9–11 and P. Sautière and P. Puigdomenech, personal communication). Sequences are taken from the following: goose H5, chicken H5, rabbit H1 and calf H1, trout testis H1 and sea urchin sperm H1 (*Parechinus angulosus*) (refs 3, 4, 1, 2 and 5 respectively). The sea urchin embryo H1 (*Psammechinus miliaris*) sequence is from a cloned DNA fragment⁷. *b*, Peptides and modified protein from calf thymus histone H1 and chicken erythrocyte H5. Calf thymus H1 was extracted with 5% perchloric acid and purified on Amberlite CG-50. It was iodinated at Tyr 72 by the addition of iodine/potassium iodide at a 1:1 molar ratio of iodine to protein in 15 mM bicarbonate buffer pH 9. The degree of iodination was determined from the NMR spectrum of tyrosine in the disordered protein and found to be 0.95 (ref. 27). NMR was also used to show that the sample of modified H1 (M-H1) does not fold²⁷. The introduction of a very bulky atom into a buried residue³⁵ of a relatively small domain is presumed to prevent folding. The peptide N-H1 was prepared by thrombin digestion of H1 (ref. 36) and purified on CM-cellulose. Identification of the fragment was by amino acid analysis (showing the absence of phenylalanine and the presence of tyrosine), by sizing on the gel (slightly smaller than histone H4 of 102 residues) and by observing the presence of the N-terminal acetyl group in the NMR spectrum. The precise C-terminal point is not established but is ± 5 residues. The NMR spectrum of the peptide at high ionic strength indicated that it did not form a folded structure³⁶. The peptide NG-H1 was also prepared by thrombin digestion³⁶. The observation of perturbed peaks in the NMR spectrum of NG-H1 demonstrated that the sample used here folds to form tertiary structure. This peptide is very similar to an abundant natural degradation product of H1, previously labelled 'HMG 8' (ref. 37). We have shown by NMR spectroscopy that HMG 8 contains the folding domain of H1 and is N terminal. Peptide C-H1 was likewise prepared by thrombin digestion of H1. It does not fold³⁶. Peptide G-H1 was made by trypsin digestion of H1 in free solution⁹ and G-H5 likewise from chicken erythrocyte H5 (ref. 10). Both samples were checked by NMR for the ability to fold. Gel electrophoresis was used to check the purity of the peptides and proteins used.

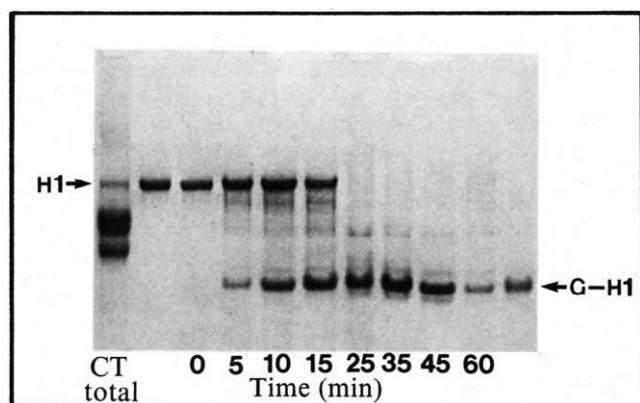


Fig. 2 Time course of trypsin digestion of nuclei. Triton X-100-washed calf thymus nuclei were suspended at 20 °C in 0.15 M NaCl, 10 mM Tris, pH 8, at a concentration of ~ 2 mg ml⁻¹ DNA. Trypsin was added to a molar ratio of 1:500 with respect to the content of total histone. Soybean trypsin inhibitor was added to aliquots at different times and extraction was with 5% perchloric acid. After precipitation with acetone, the proteins were analysed on 15% polyacrylamide-acid/urea gels³⁸ and compared with a sample of G-H1, the globular peptide obtained from free-solution digestion of H1 (ref. 9).

observed. Peptides were therefore reconstituted at values of r up to 2 as a test for their ability to condense the stripped chromatin. As reconstitution of H1 and its peptides results in a monotonic increase in S-value to a maximum of only 1.7 times that of the stripped chromatin, it is assumed that no interparticle cross-linking occurs.

The ability of peptides to compact the nucleofilament has also been tested by observing the rate at which micrococcal nuclease digests the linker DNA between core particles. It is known that removal of H1, with consequent breakdown of the higher-order structure, leads to an 8–10-fold increase in this rate¹⁶ and H1 peptides have therefore been tested for their ability to restore the protection of linker DNA against digestion.

The location of H1 has been tested by nuclease digestion to the level of the core particle monomer. It has been proposed that the primary H1 binding site lies between 140 and 160 base pairs (the best current estimates are 146 and 165 base pairs^{20–22}) as there is a pause in nuclease digestion at 165 base pairs and most of the H1 is lost only on subsequent trimming to 146 base pairs^{16,23}. The best 'assay' at present for the correct location of H1 is thus the protection of 165 base pairs of DNA in the so-called chromatosome^{22,24,25}. Although digestion of native chromatin with staphylococcal nuclease does not give 100% yields of the chromatosome before 146-base pair core particles are produced, nevertheless, if H1 is absent, no protection of 165 base pairs is observed and digestion to 146 base pairs is very much quicker¹⁶. Several peptides from H1 and H5 have therefore been reconstituted onto chromatin stripped of lysine-rich histones, and micrococcal nuclease digestion used to see whether protection of 165 base pairs is observed at a level comparable to that found in native chromatin.

Sedimentation studies

Figure 1 shows the positions of the peptides used in the intact chains of H1 and H5. Figure 3 gives sedimentation coefficients of the reconstituted chromatins relative to the stripped chromatin used for each experiment. The results can be summarized as follows: peptide N-H1 (1-98) has no influence on the sedimentation coefficient. The peptide does however bind to the stripped chromatin as it is all found in the pellet after rapid centrifugation and NMR studies have shown that the peptide H1 (1-106) binds to DNA, albeit not strongly²⁶. Thus, an N-terminal domain that lacks the folded central domain cannot condense the nucleofilament. Both the globular fragments G-H1 (36-121) and G-H5 (22-100) produce a small increase in S-value, but much below that of intact H1 or H5. If the N-terminal and globular domains are present together, as in the peptide NG-H1 (1-121), the increase in S-value is greater than for the globular domain alone but still only about half of that found for intact H1. When the other half of the molecule is reconstituted, that is, peptide C-H1 (122-213), the increase in S-value is only about one-third of that for intact H1, whereas if both peptides NG-H1 and C-H1 (representing the whole molecule) are reconstituted together the sedimentation increase is similar to that of either peptide alone. It is concluded that no single domain is capable of fulfilling the compacting role of H1, nor is a combination of the N-terminal and globular domains.

The C-terminal peptide, C-H1, when reconstituted for only 45 min, gives rise to an asymmetric sedimentation zone indicating a structurally heterogeneous sample, whereas after a 20-h incubation a symmetric zone was observed. The short incubation presumably results in irregular location of this peptide, although with time at 80 mM NaCl the peptide may find its correct location. Similar observations were made when the modified H1 (M-H1) was reconstituted, in that an increasingly asymmetric zone was observed as r was raised. Although at $r = 2$ the native S-value was restored, with M-H1 the asymmetric peak was of reduced area due to loss of rapidly sedimenting material. Increased incubation time, moreover, did not yield a symmetric sedimentation peak. It is concluded that nonspecific interaction occurs between M-H1 and the stripped chromatin that involves some cross-linking between nucleofilaments and also between incorrect sites on a single nucleofilament.

Digestion of linker DNA

Native chicken erythrocyte chromatin, or stripped chromatin reconstituted with intact calf thymus H1 or with H5 at $r = 2$, is about 10 times more resistant to cleavage within linker DNA than is the stripped chromatin^{16,18} (see Fig. 4a, c). Reconstitution with iodinated H1 (M-H1) resulted in linker DNA cleavage at about the same rate as in intact chromatin (Fig. 4b). In contrast, reconstitutes made from each of the peptides shown in Fig. 1 had a rate of linker digestion indistinguishable from that of stripped chromatin. It seems, therefore, that only the intact molecule can compact the nucleofilament to the degree necessary for providing linker protection.

Monomer length fragments

Figure 5a shows that when depleted chromatin is digested with 10 times the amount of nuclease used for linker digestion, the monomer particle is trimmed essentially to a uniform length of ~146 base pairs whereas if intact H1 is reconstituted an additional product of ~168 base pairs appears as with native chromatin¹⁸. If the N-terminal fragment N-H1 (1-98) is reconstituted, only a 146-base pair band is observed, whereas if NG-H1 (1-121) is used there is considerable protection of ~168 base pairs, as in the reconstitute with intact H1. This suggests that the globular domain is essential for the protection of 168 base pairs. This was confirmed by reconstitution with the globular domain alone, G-H1 (36-121); Fig. 5b shows that, at intermediate digestion times of 2-5 min, the bands corresponding to 168 and 146 base pairs are of comparable intensity, as

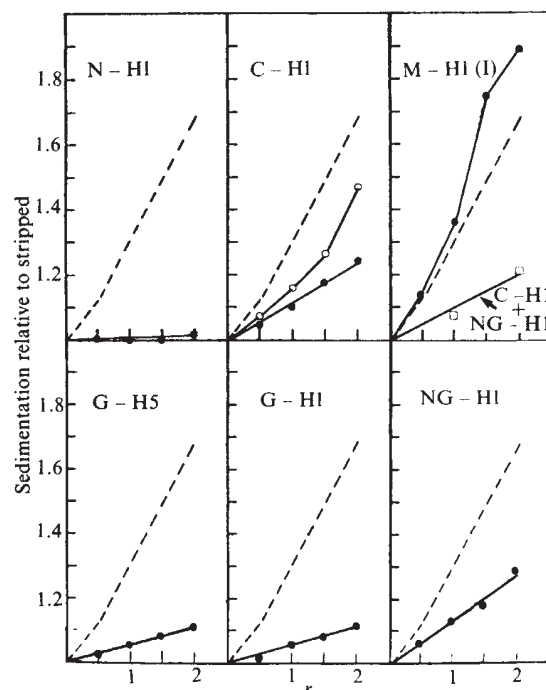


Fig. 3 Sedimentation velocity measurements of reconstituted chromatins. Chicken erythrocyte nuclei were lightly digested with micrococcal nuclease to yield chromatin of 30-50 nucleosomes length. This was depleted of H1 and H5 by passage over a DNA-cellulose column at 80 mM NaCl¹⁸. Proteins and peptides were reconstituted at several values of r (mol peptide per nucleosome) by direct addition at 80 mM NaCl. The S-values of reconstitutes were measured in 5-20% linear sucrose gradients together with native chromatin and H1/H5-stripped chromatin in separate buckets of the same rotor (about 99S and 58S respectively). S-values of reconstitutes are expressed relative to that of stripped chromatin. The broken line in each panel represents reconstitution with intact H1 (the S-value of native chromatin is reached at about $r = 2$ (ref. 19)). Reconstitutes were annealed for 45 min except for that with C-H1 which required 20 h for a reproducible result. In the experiment involving two peptides, NG-H1 was added and annealed first and the C-H1 was added subsequently.

found in reconstitutes with intact H1. The C-terminal half of H1 is unable to protect 168 base pairs, as shown by reconstituting with C-H1 (122-213); Fig. 5b shows no evidence of any band around 168 base pairs with this peptide. Thus, only peptides containing the central globular domain of H1 protect the 168-base pair fragment. To test that this domain must not only be present but must also be folded, a reconstitute was made with H1 iodinated at tyrosine residue 72. NMR spectroscopy has already shown that this modification is enough to prevent the molecule from folding²⁷ as the tyrosine is a buried residue. Figure 5b shows that very little protection of the 168-base pair fragment is afforded by M-H1—although there is some indication of a 10-base pair ladder at earlier digestion times (due probably to considerable compaction of the chromatin by this peptide), by 5 min there is no trace of DNA at the 168-base pair position.

The structural homology between H1 and H5 discussed above implies that the globular domain of H5 should also protect 168 base pairs. This was confirmed by reconstitution with this peptide, G-H5 (22-100). Figure 5b shows a strong and persistent band at ~168 base pairs throughout the digestion.

Discussion

The above data indicate that: (1) a globular domain of the lysine-rich histones (~80 residues) is protected against trypsin digestion in nuclei, and (2) this domain alone can protect the

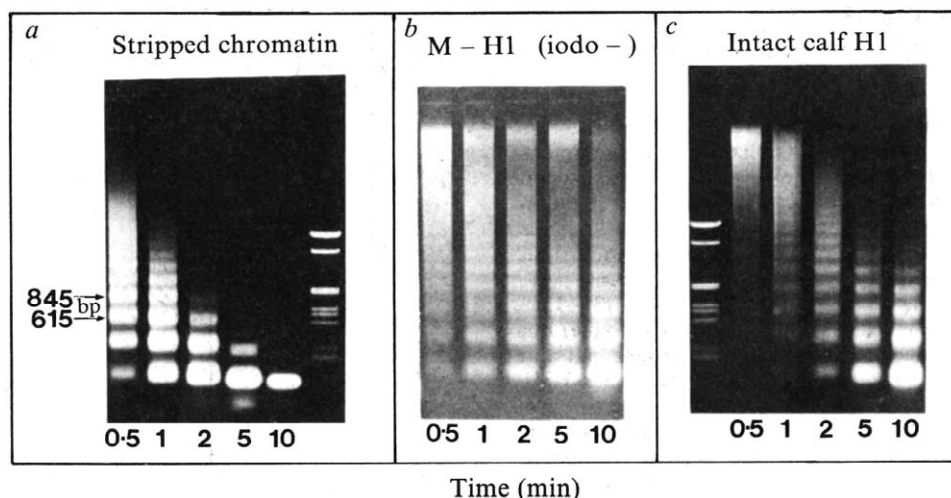


Fig. 4 Micrococcal nuclease digestion of H1/H5-stripped chromatin and reconstitutes to show preservation of the native nucleosomal repeat and changes in the rate of digestion of linker DNA. Calibration of double-stranded DNA fragments was on 1% agarose stained with ethidium bromide using a *Hae*III digest of PM2 DNA, itself calibrated with a *Taq*I digest of Φ X174 RF viral DNA. The DNA repeat length after stripping was ~ 210 base pairs (bp). Both reconstitutes were made at 80 mM NaCl and $r = 2$, and the same enzyme/DNA ratio was used in all three experiments.

extra ~ 20 base pairs of DNA present in the chromosome above that in the core particle. (3) Full compaction of chromatin by H1 may require the intact molecule.

If 146 base pairs represent $1\frac{1}{2}$ superhelical turns of DNA²⁸, then 168 base pairs is approximately two full turns. The globular domain thus seems able to close the second integral turn of DNA. 5'-end-labelling experiments have shown that the DNA extension from core particle to chromosome is symmetrical, ~ 10 base pairs on both ends (ref. 25 and L. Lutter, personal communication), and this implies a 'symmetrical' placement of the globular domain on the 2-fold axis of the particle. Figure 6 shows a scale model of this type. Crystallographic studies of the

core particle have indicated that the pitch of the DNA is ~ 28 Å (ref. 28) and this is therefore approximately the gap that would be left between the strands of DNA exiting from the particle. We have previously shown by neutron scattering¹⁰ that the diameter of the globular domain of H5 (G-H5, residues 22–100) is 29 Å. (This is very close to the value of 27 Å calculated for a sphere of 79 residues with $\bar{v} = 0.735$ ml g⁻¹.) Recent low-angle neutron-scattering measurements on G-H1 also indicate an essentially spherical particle (J. P. Baldwin and H. W. E. Rattle, personal communication). The diameter of the globular domain is therefore very close to the pitch of the DNA and is thus consistent with the model shown in Fig. 6, in which a 'cage' of

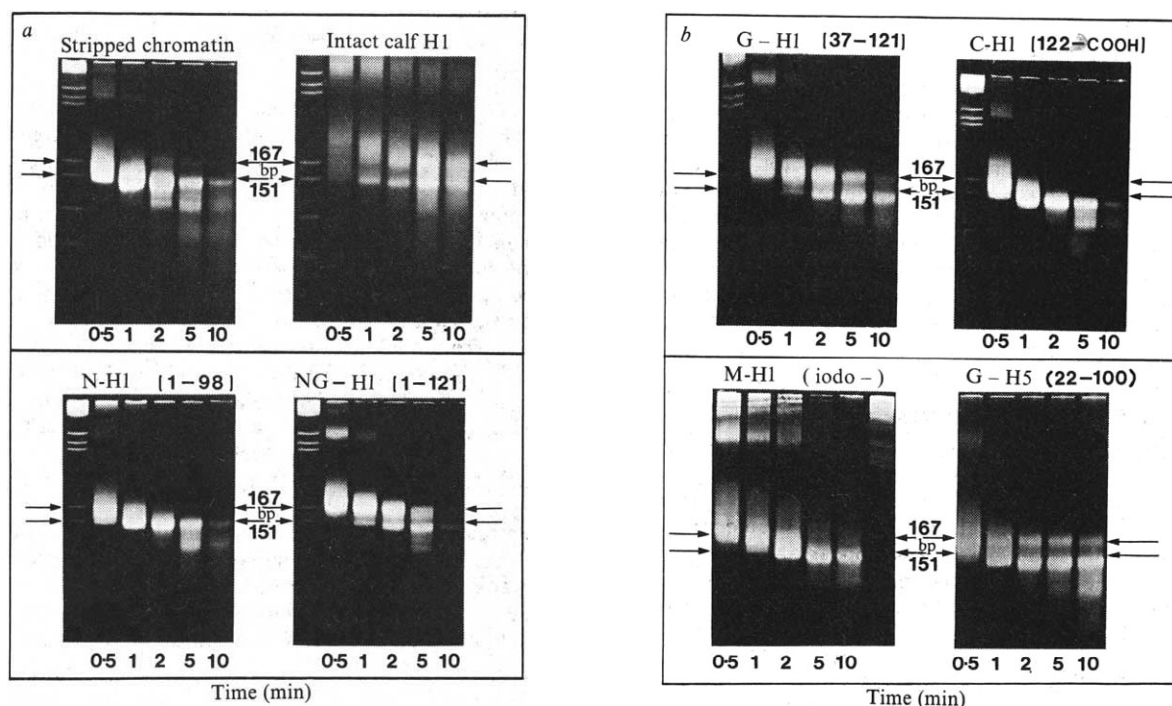


Fig. 5 Micrococcal nuclease digestion of stripped and reconstituted chromatins to study monomer-particle DNA lengths. Ten times the amount of enzyme was used as compared to Fig. 4. In all eight experiments the same enzyme to DNA ratio was used. Monomer DNA lengths were analysed on 6% polyacrylamide gels, stained with ethidium bromide. A *Hae*III digest of PM2 DNA was used as calibration. bp, Base pairs.

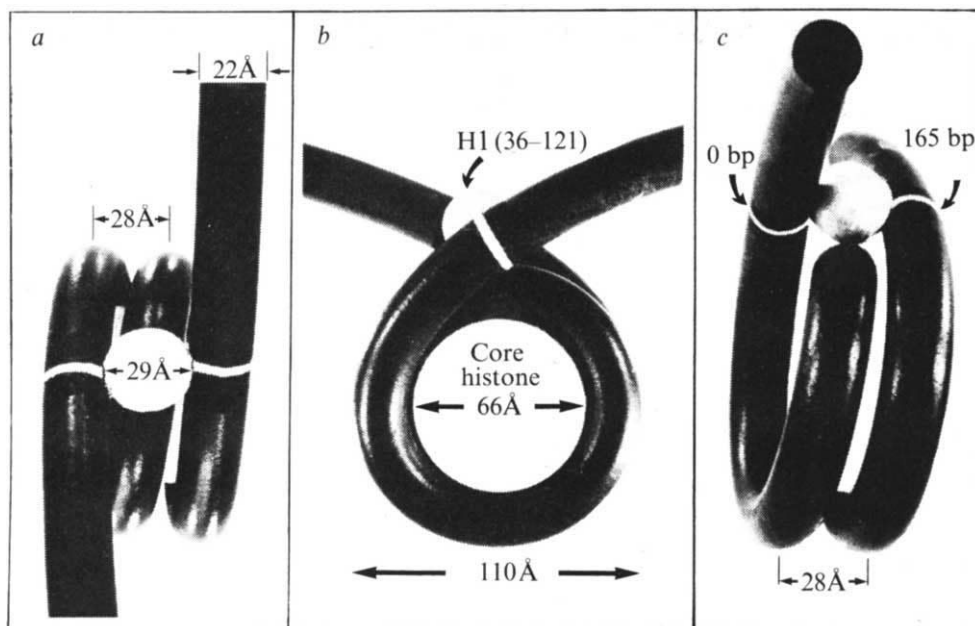


Fig. 6 Proposed model for the location of the globular domain of the lysine-rich H1-type histones at the exit point of the 165-base pair chromosome. bp, Base pairs.

three DNA strands is presented for binding of the globular domain.

This model is compatible with a number of reported observations. Singer and Singer have shown that H1 has a preference for super-helical DNA and have implicated the central region in this preference²⁹. H1 may thus bind at a point of cross-over between DNA strands. Electron microscopy has shown¹⁷ that only in the presence of H1 do the two DNA strands leave the core particle at the same point, that is, H1 is responsible for bringing the two strands together at exit point. In chromosomes there is particularly strong cross-linking between H1 and both ends of the duplex DNA, and this contact between H1 and the DNA is on the opposite side of the DNA to that at which the core histones bind; that is, H1 is bound on the outside of the particle³⁰. These data can therefore also be assimilated in the model proposed here.

The model in Fig. 6 is the most symmetrical that could be proposed and remembering the necessity for an asymmetric location of H1 (ref. 31) (and H1 is itself asymmetric), some comment is required. An asymmetric model in which 1 mole of H1 protects the extension of just 10 base pairs on one end of the core must be considered as this implies a second (equivalent) binding site giving protection of another 10 base pairs on the other end of the DNA. The existence of a second H1 binding site in the nucleosome has been proposed^{19,22} and the precise stoichiometry of lysine-rich histone in chromatin is not yet

certain^{23,31-34}. With such a model one would, however, expect that, at $r=1$, protection of predominantly a 156-base pair length would be observed, rather than ~168 base pairs, but this has not been found.

The data given above suggest that the primary function of the globular domain is to locate H1 in chromatin by closing the core particle into the chromosome. This would be expected to involve some increase in sedimentation coefficient (Fig. 3) but does not represent full condensation of the nucleofilament. Which of the other two domains is responsible for this function? As peptide NG-H1 (1-121), which contains both the locating and the N-terminal domains, does not result in the full condensation observed with intact H1, it is concluded that the N-terminal domain is not the main agent. A fundamental structural role for the N-terminal domain is also inconsistent with the considerable sequence and length variability of this domain (Fig. 1), which rather suggests that it provides a functional distinction between different H1 species. The C-terminal domain therefore remains the most likely region for the condensation of the nucleofilament and the very basic and hydrophilic sequence of this domain fit it well for extended ionic interactions with the DNA. The present data suggest, however, that on its own it cannot achieve condensation and must be correctly located in the nucleofilament by the globular domain.

This work was supported in part by a Wellcome Senior Research Fellowship to J.A.

Received 17 July; accepted 8 October 1980.

1. Cole, R. D. in *The Molecular Biology of the Mammalian Genetic Apparatus* (ed. Ts'o, P.) 93 (Elsevier, Amsterdam, 1977).
2. Macleod, A. R., Wong, N. C. & Dixon, G. H. *Eur. J. Biochem.* **78**, 281-291 (1977).
3. Yaguchi, M., Roy, C. & Seligy, V. L. *Biochem. biophys. Res. Commun.* **90**, 1400-1406 (1979).
4. Briand, G. *et al.* *FEBS Lett.* **112**, 147-151 (1980).
5. Strickland, W. N. *et al.* *Eur. J. Biochem.* **104**, 567-578 (1980).
6. Smith, B. J., Walker, J. M. & Johns, E. W. *FEBS Lett.* **112**, 42-44 (1980).
7. Schaffner, W. *et al.* *Cell* **12**, 655-671 (1978).
8. Bradbury, E. M. *et al.* *Eur. J. Biochem.* **52**, 605-613 (1975).
9. Hartman, P. G., Chapman, G. E., Moss, T. & Bradbury, E. M. *Eur. J. Biochem.* **77**, 45-51 (1977).
10. Aviles, F. J., Chapman, G. E., Kneale, G. G., Crane-Robinson, C. & Bradbury, E. M. *Eur. J. Biochem.* **88**, 363-371 (1978).
11. Puigdomenech, P., Palau, J. & Crane-Robinson, C. *Eur. J. Biochem.* **104**, 263-270 (1980).
12. Yaguchi, M., Roy, C., Dove, M. & Seligy, V. *Biochem. biophys. Res. Commun.* **76**, 100-106 (1977).
13. Littau, V. C., Burdick, C. J., Allfrey, V. G. & Mirsky, A. E. *Proc. natn. Acad. Sci. U.S.A.* **54**, 1204-1212 (1965).
14. Bradbury, E. M., Carpenter, B. G. & Rattle, H. W. E. *Nature* **241**, 123-126 (1973).
15. Billett, M. A. & Barry, J. M. *Eur. J. Biochem.* **49**, 477-484 (1974).
16. Noll, M. & Kornberg, R. D. *J. molec. Biol.* **109**, 393-404 (1977).
17. Thoma, F., Koller, Th. & Klug, A. *J. Cell Biol.* **83**, 403-426 (1979).
18. Allan, J., Staynov, D. Z. & Gould, H. *Proc. natn. Acad. Sci. U.S.A.* **77**, 885-889 (1980).

19. Allan, J. *et al.* (in preparation).
20. Lutter, L. C. *Nucleic Acids Res.* **6**, 41-56 (1979).
21. Tatchell, K. & Van Holde, K. E. *Proc. natn. Acad. Sci. U.S.A.* **75**, 3583-3587 (1978).
22. Todd, R. D. & Garrard, W. T. *J. biol. Chem.* **254**, 3074-3083 (1979).
23. Varshavsky, A. J., Bakayev, V. V. & Georgiev, G. P. *Nucleic Acids Res.* **3**, 477-492 (1976).
24. Nelson, P. P., Albright, S. C., Wiseman, J. M. & Garrard, W. T. *J. biol. Chem.* **254**, 11751-11760 (1979).
25. Simpson, R. T. *Biochemistry* **17**, 5524-5531 (1978).
26. Bradbury, E. M., Chapman, G. E., Danby, S. E., Hartman, P. G. & Riches, P. L. *Eur. J. Biochem.* **57**, 521-528 (1975).
27. Chapman, G. E., Hartman, P. G., Cary, P. D., Bradbury, E. M. & Lee, D. R. *Eur. J. Biochem.* **86**, 35-44 (1978).
28. Finch, J. T. *et al.* *Nature* **269**, 29-36 (1977).
29. Singer, D. S. & Singer, M. *Nucleic Acids Res.* **3**, 2531-2547 (1976).
30. Belyavsky, A. V., Bavykin, S. G., Goguzadze, E. G. & Mirzabekov, A. D. *J. molec. Biol.* **139**, 519-536 (1980).
31. Hayashi, K., Hofstaeter, T. & Yakuwa, N. *Biochemistry* **17**, 1880-1883 (1978).
32. Olins, A. L., Carlson, R. D., Wright, E. B. & Olins, D. E. *Nucleic Acids Res.* **3**, 3271-3291 (1976).
33. Weintraub, H. *Nucleic Acids Res.* **5**, 1179-1188 (1978).
34. Goodwin, G. H., Nicholas, R. H. & Johns, E. W. *Biochem. J.* **167**, 485-488 (1977).
35. Giancotti, V., Quadrifoglio, F., Cowgill, R. W. & Crane-Robinson, C. *Biochim. biophys. Acta* **624**, 60-65 (1980).
36. Chapman, G. E., Hartman, P. G. & Bradbury, E. M. *Eur. J. Biochem.* **61**, 69-75 (1976).
37. Goodwin, G. H., Walker, J. M. & Johns, E. W. *Biochim. biophys. Acta* **519**, 233-242 (1978).
38. Panyim, S. & Chalkley, R. *Archs Biochem. Biophys.* **130**, 337-346 (1969).

LETTERS

Photon mass at low temperature?

Joel R. Primack & Marc A. Sher

Physics Department, University of California, Santa Cruz, California 95064

The theoretical observation that gauge symmetries are generally restored at sufficiently high temperature has led us to consider here the conjecture that gauge symmetries are broken at sufficiently low temperature. The photon, for example, would then acquire a non-zero mass m_γ below some low temperature T_0 . Above this phase-transition temperature, the photon would be strictly massless. Present observational limits on the photon mass show only that m_γ is negligible for $T > 2.7 \text{ K} \sim 10^{-4} \text{ eV}$. Although we have not discovered any plausible mechanism for electromagnetic gauge-symmetry breaking and therefore cannot estimate T_0 or m_γ , there is a considerable range of experimentally accessible low temperatures for which there are no stringent constraints on m_γ . Non-zero gluon or graviton masses at low temperatures could also have observable consequences.

The effects of finite temperatures in gauge theories have been studied in some detail¹⁻³. It has been shown that, in most models^{2,4}, the gauge symmetry is restored at high temperatures, in analogy with superconductivity. In particular, in the context of grand unified theories which unify the strong, weak and electromagnetic interactions, it is now generally agreed that as we go backwards in time to the earliest moments of the big bang, as the temperature T increases, first the electro-weak $\text{SU}(2) \times \text{U}(1)$ gauge symmetry is restored for $T > T_1 \approx 10^2 \text{ GeV}$ and then the grand unified symmetry is restored for $T > T_2 \approx 10^{15} \text{ GeV}$ (refs 5, 6). Above T_1 (T_2), the weak vector bosons (grand unified leptokuark bosons) are strictly massless.

The most dramatic consequence of the conjecture that all gauge symmetries are broken at sufficiently low temperatures is the breakdown of the $\text{U}(1)$ symmetry of quantum electrodynamics (QED) at some low temperature T_0 . In a detailed analysis of both terrestrial and astrophysical observations, Goldhaber and Nieto^{7,8} deduced a very small upper limit on the photon mass, $\sim 10^{-15} \text{ eV}$. However, all of the observations to which they refer were performed in a heat bath of at least 2.7 K ($\sim 10^{-4} \text{ eV}$), the temperature of the cosmic microwave radiation; in fact, many were done at $\sim 300 \text{ K}$, room temperature. If the electromagnetic gauge symmetry were broken at a temperature T_0 , then any experiment done at a higher temperature would be done in a $\text{U}(1)$ -symmetric vacuum, and the photon would be exactly massless. Thus, all of the observations considered in ref. 7 only show that $\text{U}(1)$ is an unbroken symmetry above 2.7 K —that is $T_0 \leq 10^{-4} \text{ eV}$.

We know of no physical mechanism through which the electromagnetic gauge invariance could break down and the photon thereby acquire a mass. Only two mechanisms for breaking a gauge symmetry in a renormalizable way have been discovered: spontaneous and dynamical symmetry breaking. In spontaneous symmetry breaking, exemplified in solid-state physics by the Landau-Ginsburg theory of superconductivity, the photon would acquire a mass through coupling to a charged Higgs scalar field ϕ^* which develops a non-zero vacuum expectation value. However, as shown in ref. 3, because the Higgs field self-coupling λ must be ≤ 1 for perturbation theory to be valid, this implies that $m_\phi \leq 10^{-1} \text{ eV}$ for $T \sim 300 \text{ K}$. This is certainly inconsistent with observation, because light charged particles would be pair-produced copiously. Although heavy Higgs particles with very strong self-interactions could conceivably give rise to a phase transition in QED at low temperature, this is

not amenable to perturbative analysis nor, we believe, is it plausible.

The photon cannot acquire a mass through dynamical symmetry breaking. Dynamical symmetry breaking is exemplified in non-relativistic solid-state physics by the BCS theory of superconductivity, in which the photon acquires a mass (Meissner effect) through coupling to Cooper pairs. In relativistic dynamical symmetry breaking⁹, the gauge particle acquires a mass through its coupling to a bound state of a fermion-antifermion pair. This will not work for the photon because of charge conjugation (C). A Higgs-type scalar bound state of a fermion-antifermion pair must have even C and thus could not be absorbed by the odd- C photon to give the photon mass—unless C is broken in the phase transition. Stern¹⁰ has also pointed out that in a massless field theory in which the coupling constant is within the domain of attraction of an IR-stable origin, there is no dynamical mass generation. Although this condition is met by conventional massless QED, its extension to finite temperature QED is not clear.

Although neither spontaneous nor dynamical symmetry breaking can give rise to a photon mass at low temperature, we regard the conjecture that gauge symmetries are broken at low temperature as sufficiently interesting to warrant experimental test, especially in view of the relative ease of such tests. The main drawback of the lack of an explicit mechanism is that we cannot estimate T_0 or the photon mass, m_γ . The only insight we can offer on T_0 is the observation that $T \sim 1 \text{ K} \sim 10^{-4} \text{ eV}$ is about as far below the electro-weak phase transition ($T_1 \sim 10^{11} \text{ eV}$) as the grand unified phase transition ($T_2 \sim 10^{24} \text{ eV}$) is above it—in the logarithmic sense, relevant for the renormalization group analyses which determine T_1 and T_2 . Both symmetry-breaking models and physical plausibility suggest that $m_\gamma \sim T_0$ (T_0 is the only mass scale available). Lacking a specific model, we can offer no insight into the order of the phase transition; in a first (second)-order transition, the photon mass will change discontinuously (continuously) from zero as the system is cooled.

There would be many observable consequences if $\text{U}(1)_{\text{em}}$ were broken below T_0 and the photon acquired a mass m_γ ; we will mention two. The Poisson equation for the electrostatic potential V acquires an extra mass term

$$(\nabla^2 + m_\gamma^2)V(\mathbf{r}) = -\rho(\mathbf{r}).$$

Then the potential difference between a charged conducting sphere and an inner conducting sphere insulated from the outer one will be⁷

$$\frac{\Delta V}{V} = \frac{1}{6} m_\gamma^2 (a^2 - b^2) + O(m_\gamma^4 a^4)$$

where a and b are the radii of the outer and inner spheres. (Generally⁷ the effect of m_γ on a system of size a is $O(m_\gamma a)^2$.) This effect could be detected by connecting the inner and outer spheres briefly to allow charge to flow while the system is at low temperature, and later measuring the charge on the inner sphere. A very accurate measurement of ΔV would set stringent limits on m_γ , but would not be necessary to detect $m_\gamma \sim T_0$. A second way of detecting m_γ would be to measure the resonant frequencies of a superconducting cavity as a function of temperature.

Another consequence of our conjecture is that the $\text{SU}(3)$ gauge symmetry of the strong interactions, quantum chromodynamics (QCD), is broken at low temperature. Is it possible that Fairbank *et al.*¹¹ have been the only group to detect fractionally charged particles—quarks?—because theirs has been the only low-temperature search (N. Deshpande, personal communication)?

De Rujula *et al.*¹² have devised an MIT-bag model of broken QCD in which the eight $\text{SU}(3)$ gluons get equal mass μ , and individual free quarks would have a mass $M \sim (2\pi\alpha'\mu)^{-1}$ where

$\alpha' = 0.88 \text{ GeV}^{-2}$ is the Regge slope. Their model involves three triplets of coloured Higgs scalars, and μ is a function of the parameters of the Higgs self-coupling potential, including the effects of gluon loops^{12,13}. In view of the experimental lower limit on M of several GeV, they must consider $\mu < 10 \text{ MeV}$. (They do not discuss the behaviour of their theory as a function of temperature, although the phase transition restoring the colour gauge symmetry would be at temperatures that might be reached in extreme stellar states.)

In the model of De Rujula *et al.* there seems to be no reason why μ could not be very small, say $\sim 10^{-4} \text{ eV}$. This would exemplify our conjecture: the SU(3) colour gauge symmetry would then be broken at low temperature but restored at high temperature. However, the physical consequences would at first sight seem to be negligible: although free quarks could exist, they would have large mass because of the large size $V \sim \mu^{-1}$ of their associated colour-confining bags. The only way that light, fractionally charged particles could exist in this model would be apparently for quarks to form colourless bound states with the colour-triplet Higgs scalars. But then we do not know why this would not also happen in the unbroken phase. In any case, this is only a model—perhaps colour symmetry is broken dynamically rather than spontaneously at low temperature. As with QED, in the absence of definitive theoretical analysis we regard the low-temperature behaviour of QCD as uncertain and subject to experimental test. Other possible effects to look for at low temperature besides free quarks are slight shifts in nuclear energy levels. Unlike QED, the effects of QCD symmetry breaking at low temperature might be extremely subtle or even essentially undetectable.

Does our conjecture apply to gravitation? Although the graviton is in a limited sense a gauge particle, with the Lorentz group as the gauge group, it does not (except in non-einsteinian formulations^{14,15}) have Yang–Mills dynamics. While several authors^{16,17} have recently considered spontaneous symmetry breaking in the context of gravitation, their objective has been to obtain the gravitational coupling constant G through a scalar field with a nonvanishing vacuum expectation value, rather than to give the graviton a mass at low temperature. The possibility of the graviton having a non-zero rest mass at 2.7 K has been ruled out¹⁸. It would be possible in principle to observe a graviton mass experimentally, for example, by measuring the force at low temperature on a mass suspended between two unequal masses at a distance such that the force would vanish if the force law were exactly r^{-2} . If the graviton acquired a mass below some temperature which the universe eventually reaches as it expands, the cosmological consequences would obviously be profound.

One major difficulty with detecting a QCD or gravitational phase transition at low temperature should be emphasized. For a QCD phase transition to occur, it would be necessary for the coloured particles to come to thermal equilibrium with the heat bath. It is not clear that this could happen on a reasonable time scale. Because the gravitational coupling is so weak, it is even more doubtful that equilibrium could be established even on a cosmological time scale. In the case of QED, however, the presence of, for example, a charged Higgs field would lead to rapid equilibration with a photon heat bath.

We note that the simplest test of our conjecture is the measurement of the photon mass at low temperature. A positive result in such an experiment would have important consequences. One could study gauge theory symmetry breaking and phase transitions in the laboratory, the transfer of energy from the gauge boson rest mass to a cosmological constant and vice versa (that is, the apparent creation of energy discussed in ref. 3), a region of space with a different cosmological constant than in our Universe, and so on. We can offer no explanation for $U(1)_{\text{em}}$ breaking at a low temperature, and there is also no clue as to the value of T_0 . Nonetheless, we have proposed simple experiments which would at least lower the upper limit on the $U(1)_{\text{em}}$ -breaking phase-transition temperature by a few orders of magnitude.

We thank N. Deshpande, D. Dorfan, T. Goldman, A. Guth, P. Langacker, K. T. Mahanthappa, B. Rosenblum, D. Toussaint and D. Unger, for useful discussions. This work was supported in part by the NSF.

Received 4 September; accepted 28 October 1980.

1. Dolan, L. & Jackiw, R. *Phys. Rev. D* **9**, 3320–3341 (1974).
2. Weinberg, S. *Phys. Rev. D* **9**, 3357–3378 (1974).
3. Linde, A. D. *Phys. Rep.* **42**, 389–435 (1979).
4. Mohapatra, R. N. & Senjanovic, G. *Phys. Rev. Lett.* **42**, 1651–1654 (1979).
5. Georgi, H. & Glashow, S. *Phys. Rev. Lett.* **32**, 438–440 (1974).
6. Langacker, P. SLAC Preprint (SLAC-PUB-2544, 1980).
7. Goldhaber, A. S. & Nieto, M. N. *Rev. Mod. Phys.* **43**, 277–296 (1971).
8. Davis, L., Goldhaber, A. S. & Nieto, M. N. *Phys. Rev. Lett.* **35**, 1402–1405 (1975).
9. Pagels, H. *Phys. Rev. D* **21**, 2336–2347 (1980).
10. Stern, R. *Phys. Rev. D* **14**, 2081–2089 (1976).
11. LaRue, G. S., Fairbank, W. M. & Phillips, J. D. *Phys. Rev. Lett.* **42**, 142–145 (1979).
12. DeRujula, A., Giles, R. C. & Jaffe, R. L. *Phys. Rev. D* **17**, 285–301 (1978); **D22**, 227–228 (1980).
13. Georgi, H. *Phys. Rev. D* **22**, 225–226 (1980).
14. Yang, C. N. *Phys. Rev. Lett.* **33**, 445–447 (1974).
15. Fairchild, E. *Phys. Rev. D* **16**, 2438–2447 (1977).
16. Zee, A. *Phys. Rev. Lett.* **42**, 417–421 (1979); **44**, 703–706 (1980).
17. Smolin, L. *Nucl. Phys. B* **160**, 253–268 (1979).
18. Ford, L. H. & Van Dam, H. *Nucl. Phys. B* **169**, 126–136 (1980).

A study of PSR1237+25 at 430 MHz

T. H. Hankins

National Astronomy and Ionosphere Center, Arecibo Observatory, Arecibo, Puerto Rico 00612

G. A. E. Wright

Sonderforschungsbereich: Radioastronomie, Astronomische Institute der Universität Bonn, Auf dem Hügel 71, FRG

PSR1237+25, one of the first pulsars to be discovered^{1,2}, is characterized by a striking five-peaked integrated intensity profile. Like many other pulsars, PSR1237+25 also exhibits the puzzling phenomena of drifting subpulses, ‘nulling’ and ‘mode-switching’. The period of pulsation is 1.38 s, but the radio-emission occurs within an 80 ms interval. We have analysed single-pulse data from this pulsar at 430 MHz to throw light on the still obscure mechanism of pulsar emission. We conclude that the behaviour of PSR1237+25 can be explained on the polar-cap model if regular short (S) bursts spiralling in towards the magnetic pole are occasionally interrupted by long (L) bursts along the magnetic axis.

PSR1237+25 is nearby (at an estimated distance of 300 pc) and also bright (with a peak flux of 100 Jy at 430 MHz). Despite the complexity of the intensity profile, the position angle remains virtually constant during each pulse. Assuming that the radiation’s plane of polarization maintains a fixed angle to the magnetic field lines, this observation is consistent with a model whereby a set of open dipolar field lines surrounding a magnetic pole is being swept perpendicularly across our line-of-sight³. Within the framework of this model the instant at which the magnetic pole is pointing directly towards us is given by the longitude at which the linear polarization disappears. We shall adopt this longitude, shown on Fig. 1 as the ‘centre’ of the profile.

The four principal intensity peaks (1, 2, 4, 5) of the profile can then be ordered in terms of their decreasing displacement, r , from the central longitude. The sequence obtained is r_1, r_5, r_2, r_4 . The intensities I_1, I_5, I_2, I_4 of the peaks follow the same sequence. Although not apparent in the time resolution of our observations (1.5 ms), component 3 may consist⁴ of two closely-separated intensity peaks 3a and 3b.

The fluctuation spectrum of PSR1237+25 reveals two distinct features⁵. The first of these is a well-defined periodicity present in the first and fifth components and is ~ 2.8 times the pulsar rotation period, P (it has been shown⁵ that the spectra of different blocks of data show this feature to split into 2 or 3 slightly separated components). The second spectral feature is confined to the third component and is less well-defined, suggesting that it is not indicative of a strict periodicity, but

rather of a more generalized modulation time scale. This time scale is $\sim 40P$, or ~ 15 times longer than the first feature. Component 2 may⁵ have a weak $2.8P$ feature, but component 4 shows no evidence of periodicity.

To see how the individual pulses give rise to these features, we have examined $>2,000$ consecutive pulses, a typical section of which are shown in Fig. 2. The pulses have been normalized to their peak intensity with the absolute value of the peak given in flux units adjacent to each pulse; this emphasizes the changing morphology of the pulses. The periodicity of approximately three rotation periods is apparent in the first and fifth components, but in general the activity in these components is out of phase, with a peak in 1 preceding that in 5 by one period⁶. There is also evidence⁶ of a lesser tendency that activity in component 2 is weak when component 1 is strong, and similarly 4 tends to be weak when 5 is strong. The result of these fluctuations is typically a cycle of approximately three periods. On average, the intensity of the first component exceeds that of the fifth component in the subsequent pulse, while in the third pulse the intensity is considerably reduced and confined to the central components. The cycle then recommences in the first component. The burst-like nature of this sequence will be referred to as an S-burst. A burst begins in the first component and subsequently subsides through the components 1–5–2–4 with the third component either quiescent or closing the burst weakly. Because the bursts last slightly less than three periods (in fact $2.8P$), we gradually see them at different phases.

The simplest interpretation of these observations is that with every rotation of the neutron star the observer's beam diametrically intersects the trajectory of the S-burst in the magnetosphere, and that this trajectory is a spiral (Fig. 3). Although our choice for its sense is arbitrary, clearly the burst must begin in such a way that it first intersects the path of the observer's beam at the longitude of the first component. It then proceeds to describe a spiral with diminishing intensity through the components 5, 2, 4 and, possibly, 3a and 3b. It is also clear from inspection of the individual pulses (see Fig. 2) that in any given pulse emission is often simultaneously present in several components, suggesting that as it moves towards the centre the burst becomes linearly extended along its trajectory.

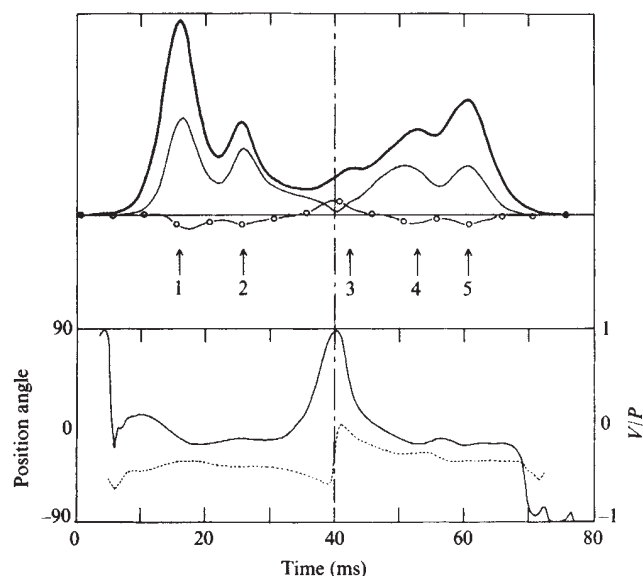


Fig. 1 Integrated polarization profile of 1,593 pulses at 430 MHz, showing the five labelled pulse components. Linear and circular polarization are shown by a light line and a line with small circles, respectively. The linear polarization angle is depicted by the dotted line and the ratio V/P where $P = (Q^2 + U^2 + V^2)^{1/2}$ is shown by the lower light line. The 'central' longitude is chosen as the point where linear polarization is a minimum, circular is a maximum, and the position angle is discontinuous.

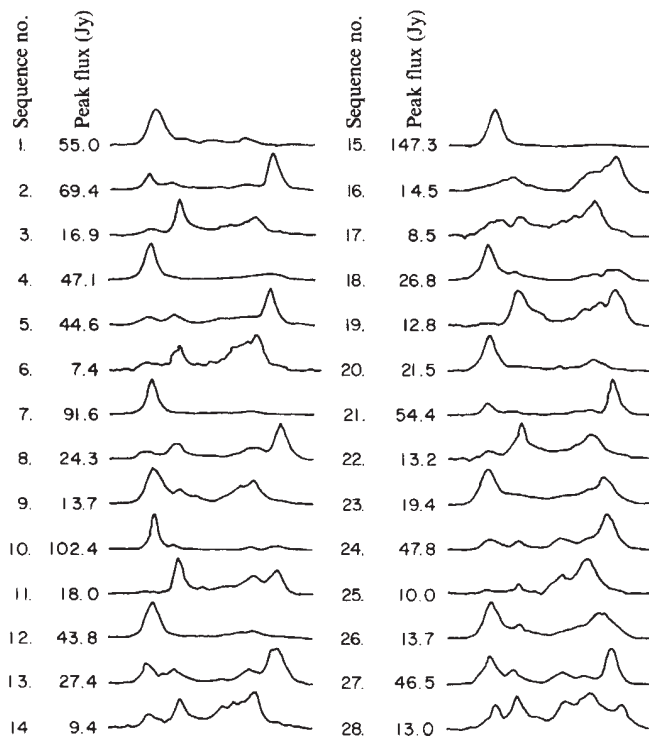


Fig. 2 28 consecutive pulses showing 10 complete S-bursts

Frequently, the burst recommences in component 1 before the previous burst has completely faded from the central components.

Interpreting the succession of the four principal components (1, 5, 2, 4) as successive increases in angular units of π ($0, \pi, 2\pi, 3\pi$), we have plotted in Fig. 4 the natural logarithm of the components' displacements from the central longitude against this angle (r being normalized by r_1). The result is approximately a straight line. This suggests that the burst moves along a logarithmic (or equiangular) spiral whose radius decreases exponentially with angle. Such spirals are described by the equation $r = r_1 \exp(-\theta \cot \alpha)$.

Two of the geometric properties of this spiral are of interest: first, its tangent (and hence the trajectory of the burst) makes a constant angle α with the radius vector. Second, the intersections of any radius vector with successive turns of the spiral form a geometric progression whose common ratio k is related to the angle α by $k = \exp(2\pi \cot \alpha)$. The straight line in Fig. 4, to which the peaks of the profile closely conform, implies a spiral with $k = 1.62$ and a corresponding angle $\alpha = 85.5^\circ$. It may only be coincidence that equiangular spirals are also formed by the shells of Nautilus Pompilius^{7,8} and other gastropods, 85° being a typical value for their determining angle.

The S-bursts do not, however, repeat indefinitely. At irregular intervals of ~ 40 pulses (or 15 S-bursts) independent activity begins close to the central longitude of the profile (in the third component). This is sometimes, though not always, accompanied by 'nulling', whereby no emission is detected above the noise level (0.5 Jy) for several successive pulses. Among the 2,100 pulses we have inspected there are 22 stretches of sustained nulls (that is, no emission was detected in two or more consecutive pulses, implying that the intrinsic lifetime of the nulls was at least a full rotation period). In every case, the pulses following the nulls showed activity in the third component and in 40% of the cases the nulls were also preceded by such activity (a third component—null association was also noted in the early work of Backer¹). These 'eruptions' in the third component will be referred to as L-bursts.

Is the $2.8P$ periodicity 'remembered' during an L-burst so that the S-bursts restart in phase? The splitting of the $2.8P$ line in the fluctuation spectrum, referred to earlier⁵, is what would be expected of the phase were 'forgotten'. The phase may be lost

during every L-burst or, less frequently, during a null sequence. The amount of splitting would, therefore, depend on the number of L-bursts or nulls included in the data blocks. It may be possible to examine this hypothesis by eliminating nulls and/or L-bursts from the data and re-determining the fluctuation spectrum. This might bring the S-bursts back into phase and make the $2.8P$ feature free of splitting and broadening. This effect has already been found¹¹ in PSR0809+74, a pulsar of similar age and period, where nulls occasionally interrupt a highly regular subpulse drift: the nulls sever the drift-bands so that only by removing them from the data is the continuity of the drift restored.

In common with PSR0329+54 and possibly other pulsars, PSR1237+25 switches occasionally into an 'abnormal' mode^{2,12-14}, whereby the emission pattern is radically changed. Only one example of this mode occurred in our data, lasting 74 pulses, but unfortunately at a time when interstellar scintillation weakened the definition of the pulses. There is considerable similarity between abnormal mode activity and that observed during the L-bursts. The mode is characterized by sustained activity in the third component and by changes in the longitude of the outer components. No S-burst periodicity can be seen, and we can again only suggest that their repetition has either become irregular or now has a time scale comparable to the rotation period. At our observing frequency (430 MHz) the components in the left hand of the profile are considerably stronger than those on the right: a feature which may be the result of longitudinally-dependent propagation effects.

The polar cap model¹⁵⁻¹⁸ is still the only model which offers a physical explanation for the drift behaviour of sub-pulses. It postulates the existence of a 100-m diameter cap surrounding the magnetic pole on the neutron star's surface, whose outer limits are determined by the last magnetic field lines to close within the light-cylinder. Above the cap a gap of height 100 m forms, in which strong electric fields are maintained both parallel and perpendicular to the magnetic field. Within this volume γ rays with sufficient photon momentum perpendicular to the magnetic field (originating either from the hot surface or

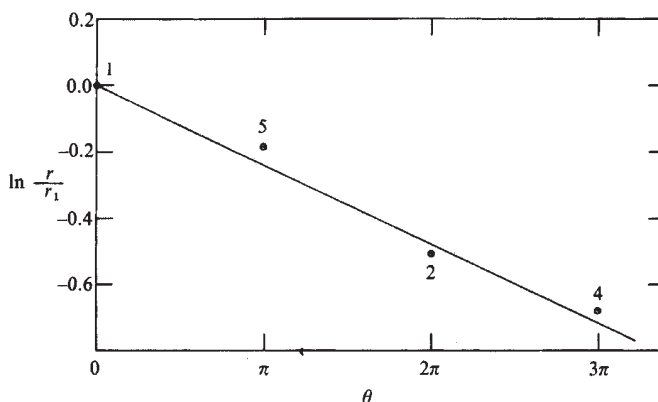


Fig. 4 Plot of $\ln(r/r_1)$ against θ . The straight line corresponds to a logarithmic spiral with a common ratio of $k = 1.62$.

from strongly-accelerated particles) generate cascades of electron-positron pairs. These particles are accelerated outwards and, following further pair-creation, excite two-stream instabilities which generate the observed coherent radiation in the upper magnetosphere (10^7 – 10^8 cm above the cap in the most recent version¹⁷ of the model).

Ruderman and Sutherland¹⁶ have suggested that the drift of sparks around the magnetic pole is the consequence of deviation in the local electric field from its 'force-free' value. The sparks are likely to arise near the edges of the cap where the field-line curvature is largest (and, therefore, where the probability of pair-creation is greatest¹⁹). Assuming the trajectories to be circles, they arrive at the expression $5.6B_{12}/P^2$ for the circulation period of the single spark region in units of P (B_{12} being the magnetic field strength in 10^{12} G). Inserting the appropriate parameters for PSR1237+25 we obtain a value of $2.9P$ for the drift period, in close agreement with the observed periodicity of $2.8P$ for the S-bursts. Ruderman and Sutherland do not discuss the possibility of non-circular drifting, but for an 85.5° spiral the inward radial velocity will only be a fraction $\cot 85.5^\circ = 0.08$ of the tangential velocity and so in a sense only a second-order effect. No obvious explanation is available for the radial drift: possibly the sparking region, having circumscribed the pole is then for some reason inhibited from repeating its trajectory and moves to an inner radius. Note that a spark following the spiral of Fig. 3 would exploit the entire surface of the cap. Alternatively, we might argue that the presence of the spark will itself distort the electric potential from its axisymmetric form (as pointed out by Jones¹⁸). There is then no reason to expect the drift trajectory to be perpendicular to the radius. It is also difficult to explain why the inner radius of the sparking region does not move to the pole within milliseconds if the cascade mechanism described by Sturrock¹⁵ is operating. Radial confinement of the spark may be more easily achieved if pair-creation is being sustained by the 'photon splash'^{17,18} of γ rays from the surface of the cap.

The simultaneous appearance of the S-bursts in several components suggests that the burst becomes linearly extended along its path. This would occur, for example, if a spark of finite angular size were to drift along a spiral with an angular velocity which increased towards its centre. Constant lateral drift velocity is one, but not the only, possibility. The angular velocity would then increase proportionately with each turn.

The behaviour of the L-bursts cannot be explained on any current theory. Their association with nulling and their modulating effect on the S-bursts suggest that an understanding of their role is crucial. Possibly the S-bursts represent charging and L-bursts discharging of the neutron star. The third component of PSR1237+25 may require a separate physical explanation^{10,14}. Perhaps this is a hint as to why in most pulsars nulls interrupt the pulse trains in such an apparently random way²⁰: if PSR1237+25 were viewed from any angle other than that which contains the magnetic axis, then the subpulse drift and the

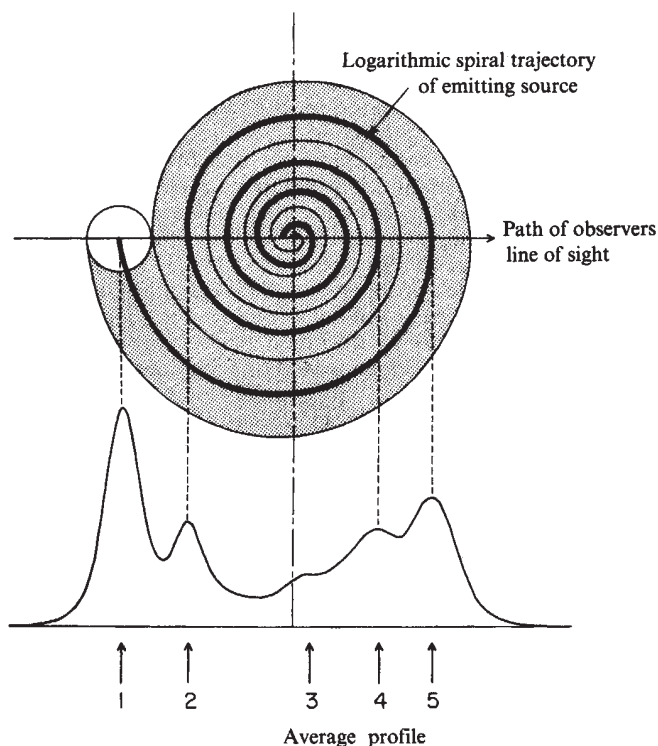


Fig. 3 The proposed spiral path of the S-burst around the magnetic pole and its relation to the observed intensity peaks. The path of the observer's line of sight is shown passing directly over the magnetic pole.

nulling would still be observed, but not the third component. Although a null involves the entire polar cap, its causal mechanism may be confined to the centre and, therefore, seldom seen. Interestingly, in several pulsars Backer⁵ has found modulations of typically 100 periods in addition to the short-term drift periodicities. Their possible connection with nulls does not seem to have been investigated.

Multiply-peaked integrated profiles are not uncommon among pulsars¹⁰, often with asymmetrically distributed peaks. PSR0329+54 and PSR2045-16 are perhaps the best and brightest examples. Intriguingly the binary pulsar²¹, whose individual pulses are too faint to detect, has also an asymmetric triple peak profile. It would, however, be rash to extend the present model to other pulsars without further investigation. The pulsars still have much to teach us, and have hidden their secrets well.

We thank N. Bartel, W. Fawley, L. Fowler, D. Morris and W. Sieber for helpful discussions, and J. M. Rankin for development of the polarization measurement and reduction system at Arecibo. G.A.E.W. thanks the Sonderforschungsbereich: Radioastronomie for support. Much of this work was done while T.H.H. was visiting Max-Planck Institute for Radioastronomy supported by a fellowship from the Alexander von Humboldt Stiftung.

Received 21 July; accepted 20 October 1980.

- Backer, D. C. *Nature* **228**, 752 (1970).
- Backer, D. C. *Nature* **228**, 1295 (1970).
- Radhakrishnan, V. & Cooke, D. J. *Astrophys. Lett.* **3**, 225 (1969).
- Backer, D. C. thesis, Cornell Univ. (1972).
- Backer, D. C. *Astrophys. J.* **182**, 245 (1973).
- Taylor, J. H., Manchester, R. N. & Huguenin, G. R. *Astrophys. J.* **195**, 513 (1975).
- Lockwood, E. H. *A Book of Curves* (Cambridge University Press, 1961).
- Thompson, D. W. *On Growth and Form* (Cambridge University Press, 1971).
- Kahn, P. G. K. & Pompea, S. M. *Nature* **275**, 606 (1978).
- Backer, D. C. *Astrophys. J.* **209**, 895 (1976).
- Unwin, S. C., Readhead, A. C. S., Wilkinson, P. N. & Ewing, M. S. *Mon. Not. R. astr. Soc.* **182**, 711 (1978).
- Lyne, A. G. *Mon. Not. R. astr. Soc.* **153**, 27P (1971).
- Morris, D., Sieber, N., Ferguson, D. C. & Bartel, N. *Astr. Astrophys.* (in the press).
- Oster, L. & Sieber, W. *Astr. Astrophys.* **58**, 303 (1977).
- Sturrock, P. A. *Astrophys. J.* **164**, 529 (1971).
- Ruderman, M. A. & Sutherland, P. *Astrophys. J.* **196**, 51 (1975).
- Cheng, A. F. & Ruderman, M. A. *Astrophys. J.* **235**, 576 (1980).
- Jones, P. B. *Mon. Not. R. astr. Soc.* **184**, 807 (1978).
- Erber, T. *Rev. Mod. Phys.* **38**, 626 (1966).
- Richtings, R. T. *Mon. Not. R. astr. Soc.* **176**, 249 (1976).
- Taylor, J. H., Fowler, L. A. & McCulloch, P. M. *Nature* **277**, 437 (1979).

Unambiguous mass determination of major stratospheric positive ions

E. Arij, D. Nevejans & J. Ingels

Belgian Institute for Space Aeronomy, Ringlaan 3, B-1180 Brussels, Belgium

There has been recent considerable interest in the stratospheric ion composition and ion chemistry, especially in view of its possible impact on aerosol formation^{1,2} and the possibility of detection of minor constituents with very low concentrations^{3,4}. The first composition measurements of positive ions in the stratosphere conducted by a rocket-borne instrument⁵, revealed the existence of the theoretically predicted proton hydrates (PH)⁶ and also of some unexpected non-proton hydrates (NPH). (Proton hydrates are primarily ions of the form $H^+(H_2O)_n$). Because low altitude rocket data may be biased by shock wave effects, such as ion fragmentation, the need for balloon flights was strongly felt. In 1978, data obtained with balloon-borne instruments⁷⁻⁹ partly confirmed and extended the rocket data. There have been several explanations^{1,5,7,8,10} of the presence of the observed NPH, but the lack of resolution of the data means an ambiguity still exists about the exact masses of these ions. We report here the first high resolution mass spectra of positive ions in the stratosphere, at 34-km altitude, where the major mass peaks can be identified unambiguously.

The data were obtained with a balloon-borne mass spectrometer, flown on 16 June 1980 with a 100,000 m³ stratospheric balloon at mid-latitude (CNES launching base at Gap-Tallard, France, 44°33'N). The instrument was launched at 20.38 UT and reached float altitude (34 km) at 22.50 UT, time at which the measurements were started. The flight lasted 8 h and thus mass spectra were obtained in both night-time and day-time conditions. Both positive and negative ions were observed. All measurements were performed at float altitude.

We concentrate here on the masses of the most abundant positive ions during night-time. The data on negative ions will be reported elsewhere.

The instrument primarily consists of a high speed cryopump, described in detail elsewhere¹¹, in which a quadrupole mass filter is built and of the associated electronics. The stratospheric ions are sampled through a small orifice (0.2 mm diameter), drilled into a thin stainless steel flange, which can be biased with respect to the gondola (draw-in potential). They are focused into the quadrupole by an electrostatic lens. After mass filtering they reach a high gain electron multiplier, the signals of which are treated by pulse counting techniques. Between the inlet aperture and the mass filter an electron impact ion source is incorporated into the lens system. It can be put on and off by remote control and enables us to analyse the residual gas in the cryopump. Simultaneously a gas mixture of argon, krypton and xenon is injected into the mass spectrometer and an in-flight mass scale calibration is performed, using the well known mass numbers of the different Ar-Kr-Xe isotopes.

The mass spectrometer covers the mass range 0-330 AMU (at high resolution) and is microprocessor controlled. This produces a very flexible system. Several modes of operation can be called on, such as positive ion sampling, negative ion sampling and the neutral or calibration mode. The mass spectra are scanned either with a constant resolution or with a constant peak width. Most of the data for positive ions were obtained with a constant peak width (Δm) of 0.9 AMU at FWHM. A higher resolution setting implies a lower transmission of the quadrupole mass filter¹². Because in the constant peak width mode the resolution ($m/\Delta m$) is increasing with increasing mass number (m), the transmission of the instrument will be strongly decreased at higher mass numbers. Therefore very few counts have been observed above mass 119 in the positive ion sampling mode.

Figures 1 and 2 show typical spectra, obtained in the mass domains 50-106 and 90-146 AMU respectively. The corresponding calibration spectra produced in flight are also shown. With such spectra, obtained in different mass domains, the major ion mass peaks could be identified unambiguously after the appropriate integration time. The observed mass peaks, fitting into the formerly observed PH and NPH schemes are summarized in Table 1 together with their identification.

Mass 60 is a minor peak and is not shown in Fig. 1. It has only been observed after long integration times in a small mass domain (53-63 AMU), but it is added to Table 1 for completeness.

From Table 1, two ion series can be clearly distinguished. The first series, containing masses 55, 73 and 91 can be identified as

Table 1 Major observed positive ion mass peaks during night-time at 34 km

Mass number (AMU)	Identification X = 41 AMU	Observed count rate (mean value counts per s)
55	$H_3O^+(H_2O)_2$	3.7
60	$HX^+(H_2O)$	
73	$H_3O^+(H_2O)_3$	20.0
78	$HX^+(H_2O)_2$	5.7
91	$H_3O^+(H_2O)_4$	1.9
96	$HX^+(H_2O)_3$	30.0
101	$HX^+X(H_2O)$	2.4
114	$HX^+(H_2O)_4$	0.6
119	$HX^+X(H_2O)_2$	3.3

PH with the general formula $\text{H}_3\text{O}^+ \cdot n\text{H}_2\text{O}$. As can be expected from the thermochemical data on water ion clustering^{13,14} mass 73 is the most abundant peak among the PH. The existence of major masses, fitting the formula $\text{NH}_4^+ \cdot n\text{NH}_3 \cdot m\text{H}_2\text{O}$ can now be excluded from the present high resolution spectra. This confirms the conclusions about NH_3 playing no significant part in the stratospheric ion chemistry at 34 km altitude⁸.

A second distinct ion sequence is the one containing masses 60, 78, 96, 114 and 119 AMU. It is generally accepted that these ions are the result of a reaction of the PH with an unknown trace gas X. As was suggested by the Heidelberg group⁸ the formula of these NPH has the general form $\text{XH}^+ \cdot n\text{X} \cdot m\text{H}_2\text{O}$. Up to now, however, there has been ambiguity about the exact mass number of X and the best information available so far was 41 ± 1 AMU.

Although Arnold *et al.*⁵ at first proposed that formaldehyde was responsible for the presence of NPH in the stratosphere, this theory was rejected after laboratory measurements¹⁵. More recently, these authors⁸ concluded from the measured relative abundances of the PH and NPH that X must have a number density of the order of 10^5 cm^{-3} at 36 km and a proton affinity larger than $175 \text{ kcal mol}^{-1}$. Consequently they suggested CH_3CN (acetonitrile) as a possible candidate for X.

The fact that no formation mechanism for CH_3CN in the stratosphere was known led Ferguson to the NaOH proposal¹. The existence of the sodium layer in the atmosphere¹⁶ made this hypothesis quite plausible. The proposal even received support from model calculations¹⁷ and laboratory work¹⁸. Recently, however, attention has been drawn to some loss mechanisms for NaOH in the stratosphere¹⁰ and MgOH was suggested as a new candidate for X. However, the exact mass, as well as the identity of X has remained an open question.

The most important conclusion, which can be drawn immediately from our data, is that the mass number of the so far unknown molecule X is 41 AMU. Although mass 42 (HX^+) has not been observed, the presence of mass 101 and 119 can only be explained by putting the mass of $\text{X} = 41$ AMU. The conclusion rules out the hypothesis of X being NaOH (at least during night-time). The other suggestion that X might be solely MgOH is very unlikely. Actually if this were the case, one should

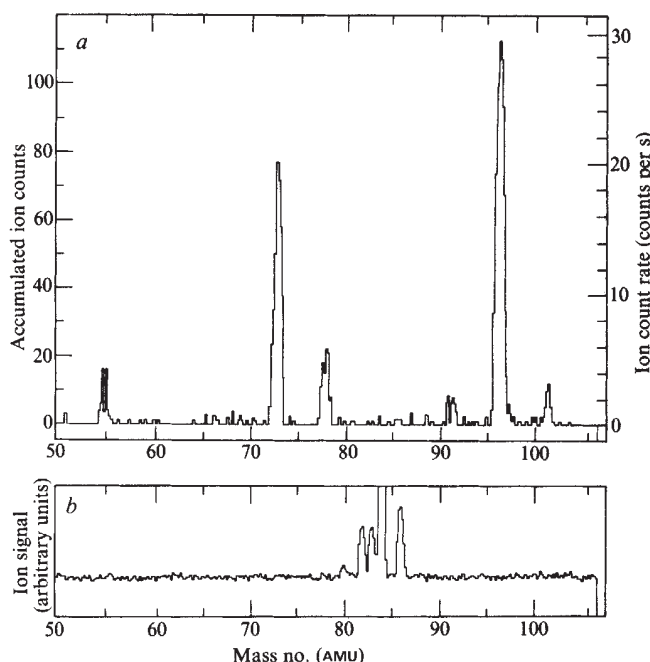


Fig. 1 *a*, Typical pre-sunrise spectrum obtained in the mass region 50–106 AMU after 14 scans. Each mass unit is divided into six channels. Dwell time per channel, 0.25 s. Duration of one scan, 84 s; $\Delta m = 0.9$ AMU FWHM. *b*, In-flight calibration obtained in the same mass domain with krypton isotopes. Only one scan is performed here.

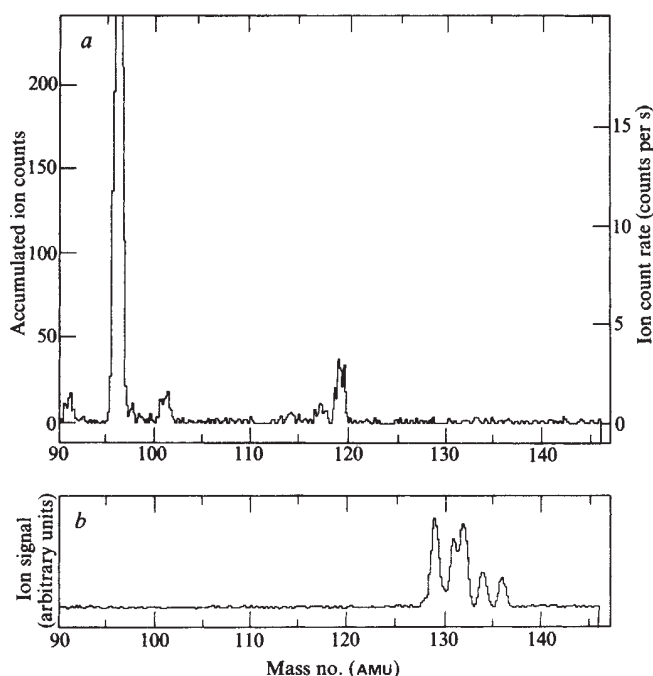


Fig. 2 *a*, Night-time spectrum in the mass region 90–146 AMU. Number of scans is 42. Programmed $\Delta m = 0.9$ AMU. Real Δm is larger in this domain. *b*, Mass scale calibration performed in flight in the same mass domain with xenon isotopes. Result of one scan.

measure a count rate at masses 97 and 98 of ~13% of that observed at mass 96, due to the isotopic abundances of ^{25}Mg and ^{26}Mg . This is, as Fig. 1 shows, not the case. Nevertheless, the possibility of a small contribution of MgOH to the NPH peaks cannot be ruled out completely; this will have to await high resolution data with large accumulated ion count numbers.

Most of the NPH peaks originate from a molecule with mass 41, different from MgOH. So far the most likely candidate for this molecule seems to be acetonitrile (CH_3CN), which has a higher proton affinity than water (187.4 kcal compared with 168.9 kcal for H_2O) (ref. 19) as well as a higher dipole moment (3.92 db compared with 1.85 db for H_2O) (ref. 20). Although no production mechanism is known for CH_3CN in the stratosphere only a very low concentration (10^5 cm^{-3}) is required to explain the conversion from PH to NPH. In this framework a critical investigation of some neutral reactions, such as those of active nitrogen (especially in excited states), which can be produced by cosmic rays²¹, with CH_4 , as well as ion molecule reactions of N^+ and N_3^+ with CH_4 (ref. 22) might be useful to look for a possible source of CH_3CN in the stratosphere. Furthermore the possibility that CH_3CN (which has been observed in interstellar clouds)²³ originates from cosmic dust cannot be overlooked.

Note that no reaction rate coefficients for the ion molecule reaction of PH with CH_3CN have been published. Although they are exothermal and are, therefore, expected to have a large rate coefficient, the study of such reactions at the appropriate temperature is strongly recommended. Up to now, however, the only certain fact about X is its mass number and even the identification of X as acetonitrile remains speculative.

Apart from the masses listed in Table 1, a peak is observed in Fig. 2, which corresponds to 117 AMU. Although this mass can be explained by the ion $\text{NaOH}_2^+ \cdot n\text{NaOH} \cdot 2\text{H}_2\text{O}$, it is very unlikely that it originates from NaOH because at night the other peaks of the series $\text{NaOH}_2^+ \cdot n\text{NaOH} \cdot m\text{H}_2\text{O}$ are not observed. It is, therefore, more probable that this peak belongs to a third series $\text{YH}^+ \cdot n\text{Y} \cdot m\text{H}_2\text{O}$. The analysis of the minor peaks indicates the existence of other members of this family. However, a more detailed analysis of the present data, as well as high resolution mass spectra with long integration times are needed to identify this series.

Finally, note that the present data are only representative of the night-time ion chemistry at 34-km altitude and that more *in situ* measurements are needed to generalize the conclusions.

We thank the technical staff of the Belgian Institute for Aeronomy and the CNES for the enthusiastic cooperation during the realization of the instrument and the balloon flight. We are also indebted to the NFWO (Belgian National Science Foundation) for partly financing this project.

Received 26 August; accepted 27 October 1980.

1. Ferguson, E. E. *Geophys. Res. Lett.* **5**, 1035–1038 (1978).
2. Arnold, F. *Nature* **284**, 610–611 (1980).
3. Arnold, F. & Fabian, R. *Nature* **283**, 55–57 (1980).
4. Arnold, F., Fabian, R., Henschen, G. & Joos, W. *Planet. Space Sci.* **28**, 681–685 (1980).
5. Arnold, F., Krankowsky, D. & Marien, K. H. *Nature* **267**, 30–32 (1977).
6. Ferguson, E. E. in *The Natural Stratosphere of 1974* (CIAP Monogr. **1**, 5.42–5.54, 1974).
7. Arijis, E., Ingels, J. & Nevejans, D. *Nature* **271**, 642–644 (1978).
8. Arnold, F., Bohringer, H. & Henschen, G. *Geophys. Res. Lett.* **5**, 654–656 (1978).
9. Olsen, J. R. *et al. Rev. Sci. Instrum.* **49**, 643–649 (1978).
10. Murad, E. & Swider, W. *Geophys. Res. Lett.* **6**, 929–932 (1979).
11. Ingels, J., Arijis, E., Nevejans, D., Forth, H. J. & Schaefer, G. *Rev. Sci. Instrum.* **49**, 782–784 (1978).
12. Dawson, P. H. *Quadrupole Mass Spectrometry and its Applications*, 142 (Elsevier, Amsterdam, 1976).
13. Kebabian, P., Searles, S. K., Zolla, A., Scarborough, J. & Arshadi, M. *J. Am. chem. Soc.* **89**, 6393–6399 (1967).
14. Mohnen, V. A. *Pure appl. Geophys.* **84**, 141–153 (1971).
15. Fehsenfeld, F. C., Dotan, I., Albritton, D. L., Howard, C. J. & Ferguson, E. E. *J. geophys. Res.* **83**, 1333–1336 (1978).
16. Megie, C., Bos, F., Blamont, J. E. & Chanin, M. L. *Planet. Space Sci.* **26**, 27–35 (1978).
17. Liu, S. C. & Reid, G. C. *Geophys. Res. Lett.* **6**, 283–286 (1979).
18. Arnold, F. in *Proc. 6th ESA-PAC Symp. on European Rocket and Balloon Programmes*, Bournemouth, 479–496 (1980).
19. Yamdagni, R. & Kebabian, P. *J. Am. chem. Soc.* **98**, 1320–1324 (1976).
20. Nelson, R. D., Lide, D. R. & Maryott, A. A. in *Handbook of Chemistry and Physics*, E-51 and E-52 (Chemical Rubber, 1972).
21. Nicolet, M. *Planet. Space Sci.* **23**, 637–647 (1975).
22. Smith, D., Adams, N. G. & Miller, T. M. *J. chem. Phys.* **69**, 308–317 (1978).
23. Solomon, P. M., Jefferts, K. B., Penzias, A. A. & Wilson, J. W. *Astrophys. J. Lett.* **168**, L107–L110 (1971).

Carbon deposition on metallic surfaces studied by r.f. plasma discharge

J. A. Cairns, J. P. Coad, E. W. T. Richards
& I. A. Stenhouse

AERE Harwell, Didcot, Oxfordshire OX11 0RA, UK

The accumulation of carbonaceous deposits on surfaces exposed to gases containing hydrocarbons or carbon monoxide is an undesirable phenomenon which can cause contamination of, for example, steel containment vessels and catalyst surfaces. It has been found¹ to occur also in certain conditions on the steel fuel pins of the Advanced Gas Cooled Reactor (AGR). Although in this instance it can be prevented by controlling the coolant composition, there is a strong motivation to understand the basic mechanisms involved. We show here, principally by means of a radio-frequency (r.f.) plasma-discharge system incorporating spectrographic analysis, that the deposits are formed largely as a result of the catalytic activities of the exposed surfaces (even here in the presence of highly ionized gases) and that deposition can be prevented by suitable coatings. These findings have now been confirmed by in-reactor experiments.

Whenever hydrocarbon molecules come into contact with metal surfaces, they exhibit a finite tendency to deposit carbonaceous material. This tendency depends on the chemical nature of the hydrocarbon. (For example, unsaturated molecules are particularly prone to such behaviour.) One example of this effect is the tendency for carbonaceous material to deposit on to the surfaces of the stainless steel fuel pins in an AGR. These fuel pins, which are manufactured from 20%Cr/25%Ni/Nb stabilized steel (20/25/Nb steel) are designed to operate within a temperature range 350–750 °C and be subjected to a coolant

which consists mainly of CO₂, but also contains significant concentrations of CO (a few per cent) and CH₄ (a few hundred p.p.m. by volume). The latter component was specified to suppress any tendency for the CO₂ to oxidize the graphite moderator². Although carbonaceous deposition can be prevented in practice by choosing specific levels of CO and CH₄ additions, it remains important to understand the basic reasons for its formation. The environment within which these events take place, namely the core of a nuclear reactor, renders direct observation very difficult, and subsequent examination of steel specimens exposed to the reactor flux is complicated by their induced radioactivity.

We have subjected specimens of the steel to an r.f. plasma discharge of a CO₂/CH₄ gas, thereby providing an irradiation environment free from the problems of induced radioactivity. It then becomes possible to compare in laboratory conditions the amounts of carbonaceous material deposited on the steel to that deposited on other surfaces, particularly copper, which has been observed³ to remain clean in reactor conditions.

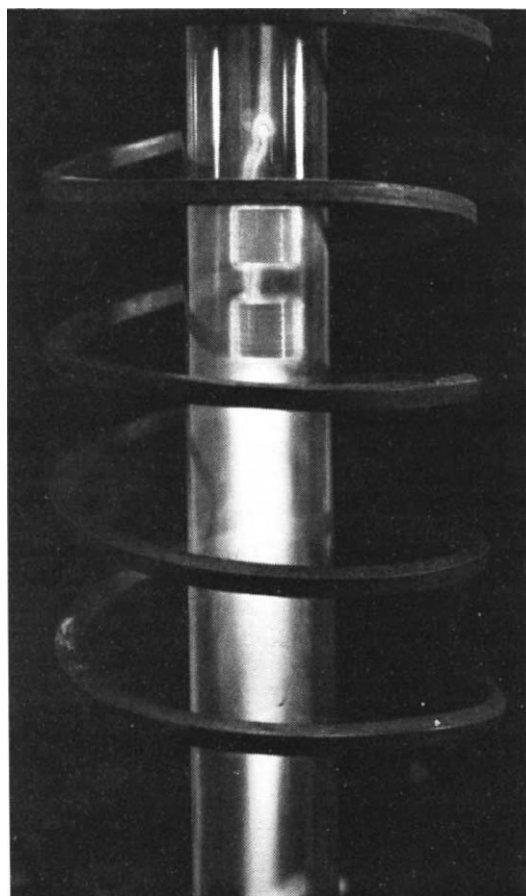


Fig. 1 Experimental arrangement showing specimens suspended in a CH₄/CO₂ r.f. plasma discharge.

Figure 1 shows the r.f. equipment in operation. Sections of AGR 20/25/Nb steel tubing (15-mm diameter) were mounted in the middle of a silica tube, surrounded by flowing gas at a pressure of 4.7 mbar and subjected to a r.f. plasma discharge (2kV; 2.5 MHz). The gas used was an equivolume CO₂/CH₄ mixture. This methane level is, of course, much higher than that present in an AGR but it was necessary here to counteract the enhanced oxidative capacity of the CO₂ induced by the intense field. In spite of this, the main effects observed previously in AGR were reproduced. Thus the steel acquired a significant deposit of carbon (0.2–0.5 mg cm⁻²) after 4 h, whereas speci-

mens of copper mounted alongside the steel remained essentially clean. Incidentally the carbon was measured by combustion in a Leco CS 144 carbon/sulphur analyser and determined as CO_2 using IR detection.

The above measurements on copper and steel indicate a clear influence of the surface chemistry of the steel in promoting deposition. To nullify this surface activity and thereby suppress the deposition, the steel was coated with a silica layer (1–2- μm thick). A convenient method is to contact the steel with a silica sol, which on heating (to 850 °C for 15 min) densifies to form a dense, protective silica coating⁴. The result was dramatic: the silica-coated steel remained clean throughout its exposure to the CO_2/CH_4 discharge, as may be confirmed from Fig. 2 which shows also that nearby uncoated steel specimens acquired carbonaceous deposits. These visual observations were supplemented by carbon analysis, which confirmed that the silica-coated specimen had acquired a negligible amount of deposit. We now consider why the silica-coated specimen remained uncontaminated: was it due to carbonaceous material being removed dynamically with enhanced efficiency throughout the exposure, or was it due to the fact that the silica surface was unable to induce deposition from its surrounding gas, even in the presence of the ionizing environment? The answer can be found by a spectroscopic examination of the light emitted from the plasma, particularly in the vicinity of the specimens.

Light emitted from a 5-cm long section of the discharge was imaged on to the slit of a Hilger medium quartz spectrograph. It was thereby possible to compare the relative emission intensities of individual spectral lines to the region from which they arose in the plasma. Identification of spectral lines and variations in these, as compared with a standard CO_2 discharge, was done using a Hilger projection comparator. A comparison between some of the species present in the vicinity of a 20/25/Nb steel surface and a corresponding silica-coated specimen, during exposure to an equivolume CO_2/CH_4 discharge, is shown in Table 1.

When the gas remote from the sample was examined for these species, it was found to contain principally CH , H and CO_2^+ , although the signal strengths were lower than in the highly ionized region near the specimens. Of particular interest are the strong signals in the vicinity of the uncoated steel of CO^+ and O_2^+ . Both of these were very weak or not detected in the bulk of the plasma or, as seen from Table 1, in the vicinity of the coated steel. Therefore it may be deduced that methane tends to suppress O_2^+ by gas-phase interaction. It seems reasonable to suggest further that methane is removed dynamically in the vicinity of the uncoated steel (hence the reappearance of O_2^+), being converted under the catalytic influence of the steel's surface to a carbonaceous deposit. This in turn is subject to attack by the oxidizing species, thereby accounting, at least partly, for the appearance there of the strong CO^+ signal. The lack of carbonaceous deposition on to the silica-coated steel specimen can be understood also: the silica surface, being free from sites which catalyse deposition, remains essentially clean and so there is little CO^+ production. The O_2^+ signal is low because the local concentration of methane is retained.

Therefore, the results suggest that carbonaceous deposition on to the steel surface is induced largely by the catalytic activity

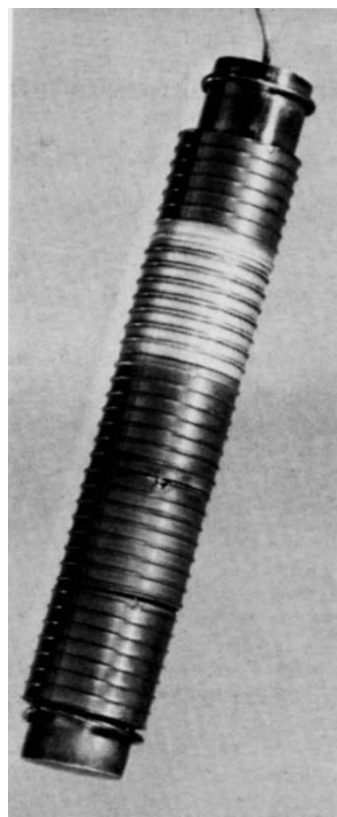


Fig. 2 The reduction in carbonaceous deposition exhibited by a silica-coated steel tube section as compared with nearby, uncoated specimens.

of the surface itself. Surfaces such as silica remain clean because they do not intervene to alter the gas composition in their immediate environment.

The r.f. plasma-discharge system used here has not only highlighted the importance of coatings to prevent carbonaceous deposition, but may be used also to screen suitable coatings for in-reactor assessment. The best coatings identified by this screening have been confirmed to perform well in nuclear reactor conditions.

Finally, there is much scientific literature on the nature of species formed as a result of ionization of CO_2 and of CO_2/CH_4 (ref. 5). However, the main difference between those studies and the present one is that here we emphasize the influence of surfaces, and particularly the 20/25/Nb steel surface in promoting carbonaceous deposition. We cannot be certain that the species identified here at low pressure will exist in that form at AGR pressures and, of course, the composition of the plasma gas examined here differs substantially from that of a genuine AGR coolant. We can conclude, however, that the success in reproducing some of the more significant gas–solid interactions indicates a general validity to the conclusions which have been reached so far.

We thank J. B. Sayers, R. H. Flowers, J. Wright, R. A. Dugdale, J. S. Hislop, C. J. Wood, D. Norfolk, P. Campion, S. J. Thomson and D. Clark for useful discussions, and P. S. Drew and C. F. Ayres for technical assistance.

Received 11 July; accepted 30 October 1980.

1. Wood, C. J. *Nucl. Engng* **22**, 29 (1977).
2. Wright, J., Dominey, D. A., Jacquiss, M. T., Linacre, J. K. & Lind, R. *Proc. 4th Int. Conf. on Peaceful Uses of Atomic Energy* **10**, 415 (1971).
3. Bennett, M. J. *et al.* in *Gas Chemistry in Nuclear Reactors and Large Industrial Plant* (ed. Dyer, A.) 81 (Heyden, London, 1980).
4. Cairns, J. A., Nelson, R. L. & Woodhead, J. L. U.K. Patent Publ. No. 2023453 A (1980).
5. Whickham, A. J., Best, J. V. & Wood, C. J. *Radiat. Phys. Chem.* **10**, 107 (1977).

Table 1 Comparison of species in the vicinity of 20/25/Nb steel with a silica-coated specimen

Species	Wavelength (nm)	20/25/Nb steel	Silica-coated 20/25/Nb steel
CO^+	219–270	Strong	Very weak
O_2^+	283.3	Strong	Not detected
CO_2^+	288.3–289.6	Very strong	Strong
CH	387.2–431.5	Very strong	Very strong
C_2	400–440	Weak	Weak
H	434–486	Strong	Weak

Generation of light hydrocarbons in sedimentary rocks

J. M. Hunt*, A. Y. Huc† & J. K. Whelan*

* Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543

† Laboratoire de Géologie Appliquée, CNRS-ERA 601, Université d'Orléans, 45046 Orléans Cedex, France

Most petroleum arises from the thermal alteration of organic matter in fine-grained sedimentary rocks. The time-temperature conditions of petroleum formation have been defined by numerous laboratory and field studies¹, but little is known of the conditions in which individual hydrocarbons form in nature. The light hydrocarbons are of particular interest because they are present only in traces in surface sediments, yet they constitute about 30% of a typical petroleum. We have found that the distribution of individual C_6 – C_8 hydrocarbons in shale samples from a Gulf of Mexico well indicates that different types of structures are being generated preferentially at different depths. Hydrocarbons with a tertiary carbon atom apparently formed earlier than those with a quaternary carbon because of the greater stability of the intermediate tertiary carbonium ion or free radical.

The samples analysed were well cuttings from a Coastal Offshore Stratigraphic Test (COST) well drilled by a consortium of oil companies about 50 miles east of South Padre Island, off south Texas in the Gulf of Mexico. The cuttings were washed, canned at the well site and kept refrigerated. At the laboratory the drilling mud was washed from the cuttings before analysis.

The C_6 – C_8 hydrocarbons were analysed by a headspace method involving the grinding of the cuttings in a closed container and release of the entrained gases in a helium headspace². The gases were analysed by capillary gas chroma-

tography on a column which gives complete resolution of all the C_6 – C_7 and part of the C_8 hydrocarbons³. Individual hydrocarbons were identified from known retention times and by gas chromatography-mass spectrometry.

The samples analysed in the 4,000–13,000 feet depth interval are Upper Miocene in age whereas the samples extending down to 16,000 feet are Lower Miocene. There are a few silts and sands in the shallow part of the well, but most of the lithology is fine-grained gray shale with organic carbon contents in the range 0.75–1.3%. An over-pressured section extends from about 10,000 feet to total depth.

Figure 1 shows the distribution of total C_6 and C_7 saturated hydrocarbons, alkanes and cycloalkanes in ng per g of sediment. The generation curve for these hydrocarbons shows a threshold of intense generation around 10,000 feet and a peak around 13,500 feet. However, the distribution of individual hydrocarbons shows a different pattern. Figure 2 shows the distributions of two dimethylbutanes and two dimethylpentanes. The differences in these structures is that the 2,2-isomers contain a single quaternary carbon atom and no tertiary carbon atoms, whereas the 2,3-isomers contain two tertiary carbon atoms. The latter structure for both the substituted butane and pentane shows a strong increase in concentration around 10,000 feet. At this depth the subsurface temperature is $\sim 115^\circ\text{C}$. In contrast, the structures with a quaternary carbon atom show slight increase at 12,000 feet with the major increase around 13,500 feet equal to a subsurface temperature of about 150°C .

Figure 3 shows the hydrocarbon distributions for a C_7 branched alkane and a cycloalkane both containing a tertiary carbon atom. They are compared with a C_8 branched alkane containing a quaternary carbon atom. Again, the latter compound shows no significant increase in concentration before reaching the 13,500-feet depth range. The compounds with the tertiary carbon atoms show this increase at 10,000 feet.

Figure 4 shows two dimethylcyclopentanes in which the only difference is the position of the methyl groups on the ring. In the 1,1-isomer there is a quaternary carbon atom whereas in the 1,2-isomer there are two tertiary carbon atoms. The latter shows increased concentrations at 9,500 feet, the former at $\sim 11,500$ feet.

The distribution patterns of the C_6 – C_8 hydrocarbons in this well show that there is a significant difference in the time-temperature requirements for the formation of hydrocarbons of different structure. These differences occur over the depth interval from $\sim 10,000$ to $\sim 13,500$ feet. Deposition rates for the Miocene in this well were $\sim 1,000$ feet per Myr, indicating a time differential of ~ 3.5 Myr and, as previously indicated, a temperature difference of 35°C for the initiation of intense generation of these different hydrocarbons.

There are several possible causes of the observed distributions: (1) a change in organic source material with depth; (2) differential migration; and (3) a change in the generation mechanism as a function of increasing depth (temperature). A change of organic source material with depth can be ruled out by the similarity of capillary gas chromatography analyses carried out on pyrolysis products generated from the organic material⁴. Moreover, migration cannot be the explanation. Light hydrocarbons in this range diffuse from their point of origin towards the surface⁵ in such a way as to cause a change in hydrocarbon concentration with molecular size because of differences in diffusion coefficients—for example, between *n*-propane and *n*-hexane. However, for structural isomers such as those in Fig. 2, the quaternary isomers diffuse more rapidly than the tertiary. Neopentane diffuses more rapidly than isopentane, for example⁶. Thus diffusion would lead to the opposite of the observed distributions and cannot be the cause. Similarly, other possible migration mechanisms do not explain the distributions. Separation of isomers due to partial adsorption during migration, or differences in boiling point, would both cause the quaternary structures to occur at shallower depths—to migrate ahead as on a laboratory chromatographic column. Bulk transport in the gas phase might result in some density separation, but again, the

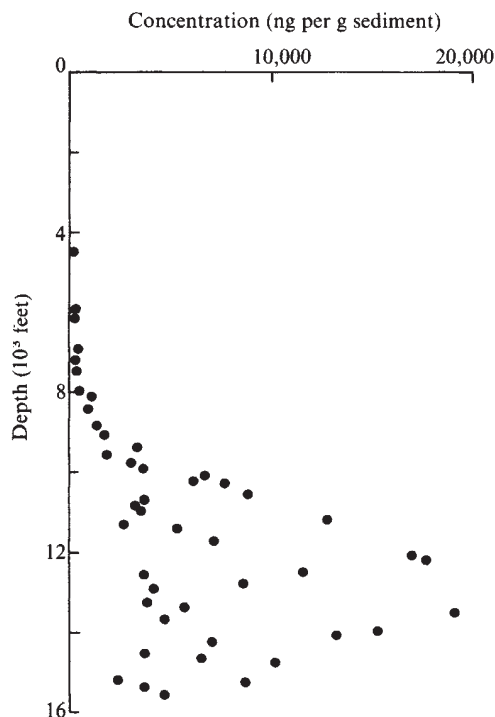
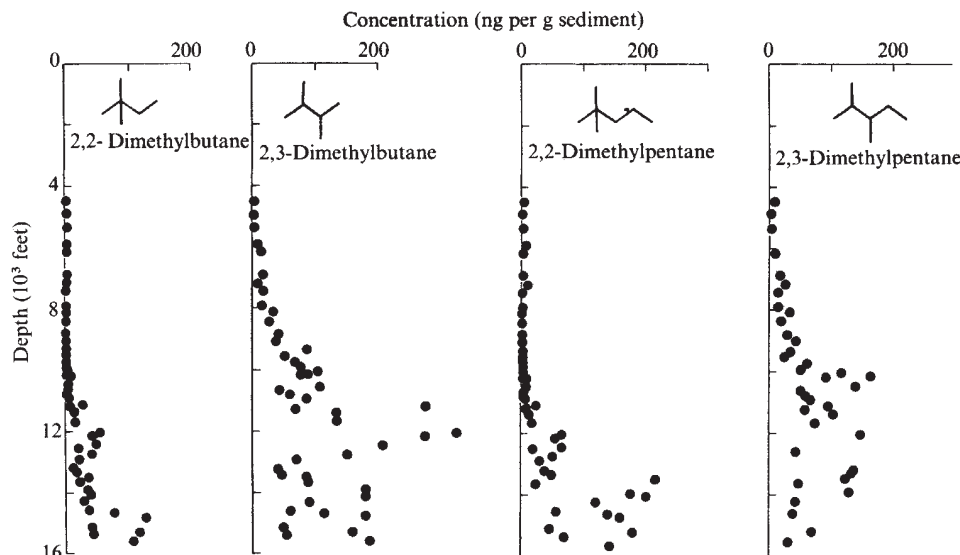


Fig. 1 Concentration plotted against depth of total C_6 plus C_7 saturated hydrocarbons in ng per g of dry sediment in the COST no. 1 well, Gulf of Mexico.

Fig. 2 Concentration plotted against depth of two C₆ and two C₇ branched alkane isomers in the COST no. 1 well.



quaternary structures are the least dense and would be found shallower instead of deeper.

These arguments suggest that the generation mechanism itself must be the explanation for the profiles observed. It seems unreasonable at these low temperatures (<150 °C) to propose thermodynamic equilibrium between isomeric saturated hydrocarbons, so that we assume that the products preferentially formed at any particular depth result from whichever intermediate (carbonium ion or free-radical) forms most rapidly (kinetic control).

There are two major generation pathways to the observed products from kerogen. Carbonium-ion cracking dominates at lower temperatures (shallower depths) and free-radical cracking dominates at higher temperatures (ref. 1, pp. 122–130). Carbonium ions rearrange whereas free radicals do not⁷. Also, tertiary carbonium ions are more stable than the primary or secondary ions and the latter easily rearrange to the former⁷. Consequently, the increased yield of tertiary structures at 10,000 feet may be the result of intensified carbonium-ion

cracking and isomerization. Carbonium ions with quaternary structures would rapidly rearrange to tertiary products, so that no increase in quaternary hydrocarbons would be observed at this depth.

Eventually, as the sediments are buried deeper to higher temperatures, free-radical cracking takes over. This causes an increase in the yield of quaternary structures at 13,500 feet because these free radicals cannot rearrange.

The fact that the yield of tertiary structures is still high at that depth suggests either that tertiary free radicals also are cracking off or that some carbonium-ion cracking is still occurring.

Whether these structures are breaking off as free radicals in the configurations shown or are isomerization products of some carbonium-ion precursors, it is significant that the hydrocarbon content of these shale source-beds is varying substantially with depth. The composition of the C₆–C₈ hydrocarbon mixture at 10,000 feet is quite different from that at 13,500 feet. This means that attempts to correlate crude oil composition with the hydrocarbons in a presumed source rock or to derive correlations between source rocks will need to account for the formation of different hydrocarbon structures in different time-temperature conditions.

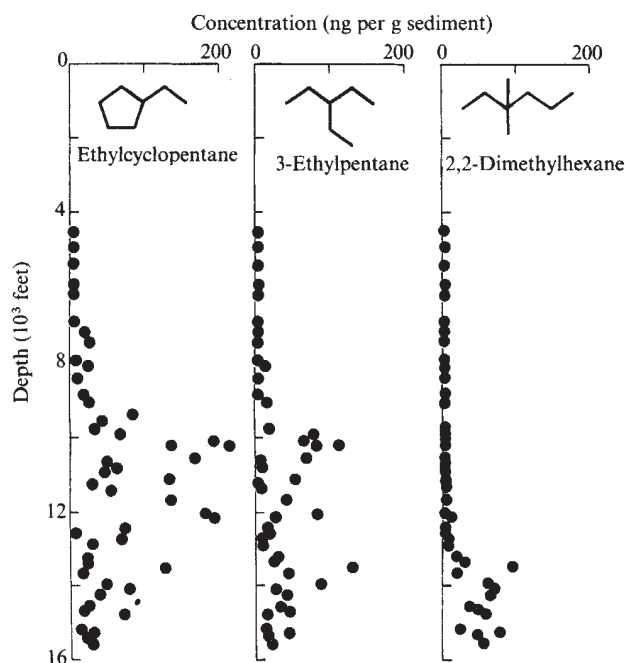


Fig. 3 Concentration plotted against depth of a C₇ cycloalkane and a C₇ and C₈ branched alkane in the COST no. 1 well.

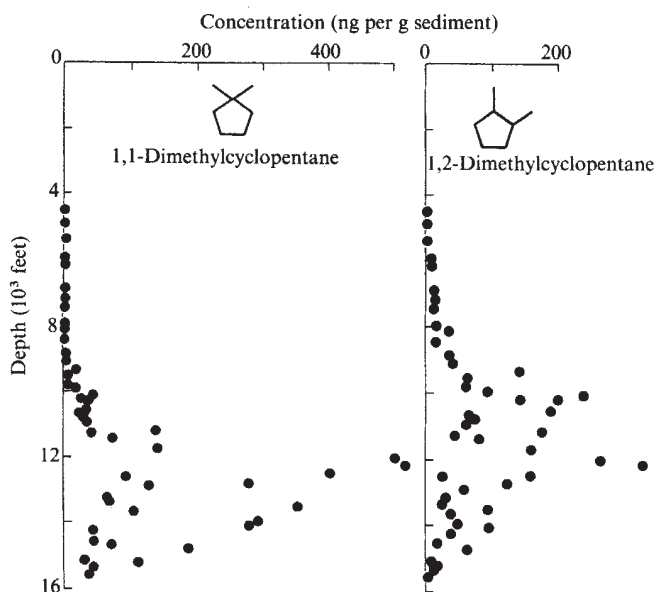


Fig. 4 Concentration plotted against depth of two dimethylcyclopentanes in the COST no. 1 well.

Individual hydrocarbons in the gasoline range (C_6 – C_8) show a significant increase in concentration at varying depths in sedimentary rocks. Hydrocarbon structures with a tertiary carbon atom increased markedly at a subsurface depth equivalent to a time and temperature of 10 Myr and 115 °C. In contrast, structures with a quaternary carbon atom did not show a substantial increase until reaching a depth equivalent to 13.5 Myr and 150 °C. The difference in depth distribution is believed to be due to the formation of carbonium ions from the kerogen at lower temperatures than free radicals. The early-forming carbonium ions then rearrange to the stable tertiary structure. Greater energy (higher temperature) is required to crack the quaternary free-radical structures which do not rearrange.

This work was supported by the Division of Basic Energy Sciences of the US Department of Energy, contract no. EG-77-S-02-4392.

Received 9 July; accepted 21 October 1980.

1. Hunt, J. M. *Petroleum Geochemistry and Geology* 131–143 (Freeman, San Francisco, 1979).
2. Whelan, J. K., Hunt, J. M. & Berman, J. *Geochim. cosmochim. Acta* (in the press).
3. Whelan, J. K. *Init. Rep. DSDP*, 47, 531–539 (1979).
4. Huc, A. Y. & Hunt, J. M. *Geochim. cosmochim. Acta* 44, 1081–1089 (1980).
5. Leythaeuser, D., Schaefer, R. G. & Yukler, A. *Nature* 284, 522–525 (1980).
6. Sahores, J. M. & Witherspoon, P. A. *Adv. Org. Geochem.* 219–230 (1966).
7. March, J. M. *Advances in Organic Chemistry* Ch. 5 (McGraw-Hill, New York, 1968).

Holocene phosphorite on the East Australian continental margin

G. W. O'Brien & H. H. Veeh

The Flinders University of South Australia, Bedford Park, South Australia 5042, Australia

Carbonate–fluorapatite, the major component of marine phosphorite, forms within organic-rich diatomaceous sediments in present day conditions along the continental margin of Peru–Chile and South-West Africa (Namibia)^{1–4}. Both of these areas are 'West-Coast' phosphogenic provinces^{5,6}, characterized by an eastern boundary current, with strong coastal upwelling and associated high organic productivity. We report here the first radiometric evidence supporting Holocene apatite formation in an 'East-Coast' phosphogenic province, the East Australian continental margin, an area of only moderate seasonal upwelling^{7,8}.

The phosphorites of the present study were recovered from the outer continental shelf and upper slope between 29 and 32° S (Fig. 1). In this region, the shelf is fairly narrow (25–40 km) and the slope is unusually steep, with the major shelf break occurring at depths^{9,10} between 210 and 450 m. A prominent feature of the oceanography of this area is the East Australian Current, considered a complex series of anti-cyclonic eddies which move generally southwards along the coast^{11,12}. Limited upwelling occurs off Evans Head and Laurieton^{7,8} (31° 39' S), causing small but significant nutrient enrichments in surface waters.

Phosphorites on the East Australian continental margin were first described by von der Borch¹³, who suggested a mid-Miocene age for them on the basis of palaeontological evidence. Preliminary uranium-series disequilibrium studies¹⁴ confirmed the relict nature of well consolidated, ferruginous phosphorite nodules from the Coffs Harbour region, but also indicated a much younger age (5.5×10^4 yr) for less well consolidated nodules from an area just south of Evans Head. We have examined phosphorites from different locations in an attempt to establish their distribution in space and time, and thus relate their formation to known environmental parameters in this area.

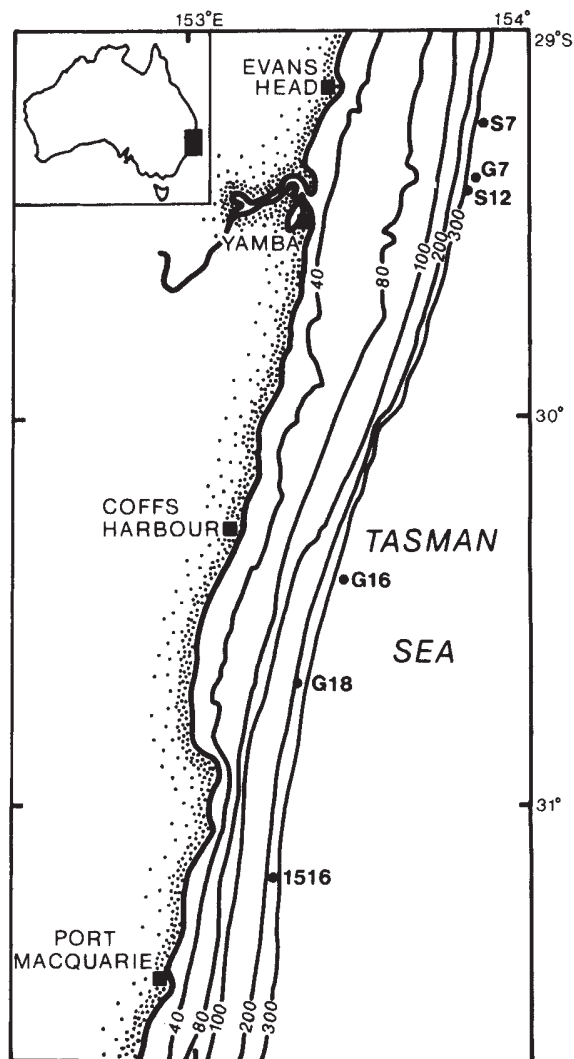


Fig. 1 Location map of phosphatic nodules on the East Australian continental margin. Bathymetric contours (in metres) redrawn from ref. 10.

The nodules fall into two easily distinguishable types—Types I and II. Type I nodules occur within a depth interval 360–420 m on the upper slope, predominantly in the Evans Head–Yamba Head area. These nodules lie within sediments rich in glauconite, biogenic carbonate and quartz and are fairly small (<4 cm diameter), earthy, friable and range in colour from near white to yellow-brown to dark grey. Microscopically, they are seen to be composed of an admixture of silt-sized angular quartz, green to green-brown rounded glauconite as well as calcareous skeletal material which has been cemented by light brownish-tinged cryptocrystalline apatite (collophane). Scanning electron microscopy (SEM, Fig. 2) shows that some phosphate is present as crystallites on the internal surfaces of foraminiferal tests, though no direct evidence for replacement of carbonate by apatite has been observed in these nodules. X-ray diffraction analysis (XRD) of Type I nodules shows quartz, high and low-magnesium calcite, aragonite, glauconite and carbonate fluorapatite as major components. Location G16 differs in that the phosphorite there is present as a poorly consolidated material infilling the interstices between the septa of solitary corals. Type II nodules, which occur off Coffs Harbour in depths of <300 m, are typically fairly large (>5 cm), highly indurated, heavily ferruginized and usually have a 'glazed' appearance. XRD indicates quartz, glauconite, goethite and carbonate fluorapatite as major constituents. Calcite and aragonite are much less abundant in Type II nodules.

Table 1 Bulk chemical composition, average collophane microprobe analyses and elemental ratios of phosphatic nodules from East Australian continental margin

Sample no. Type	Bulk chemical composition*				Average collophane microprobe analyses	
	G7-8 I	G7-9 I	G7-10 I	1516-A II	I†	II‡
Fe ₂ O ₃	4.37	5.83	5.42	27.48	2.42	8.60
MnO	0.03	0.04	0.04	0.11	—	—
TiO ₂	0.37	0.30	0.28	0.36	—	—
CaO	29.6	32.2	35.3	14.3	54.59	51.06
K ₂ O	1.05	0.97	0.93	1.31	0.17	0.11
SiO ₂	28.3	23.5	19.4	28.1	3.48	1.71
Al ₂ O ₃	4.30	3.24	2.74	4.81	1.51	0.99
P ₂ O ₅	7.5	12.2	15.3	7.2	26.62	27.38
MgO	1.38	1.35	1.24	2.51	1.27	1.21
Na ₂ O	1.3	1.2	0.9	0.4	0.47	0.77
SO ₃	0.69	0.74	0.85	0.41	1.12	1.45
F	1.9	1.8	1.9	1.3	—	—
LOI	19.5	17.6	17.1	11.7	—	—
Total	100.29	100.97	101.40	99.99	—	—
Less O	0.80	0.76	0.80	0.55	—	—
Total	99.49	100.21	100.60	99.44	—	—
CaO/P ₂ O ₅	3.95	2.64	2.31	1.99	2.05	1.86
F/P ₂ O ₅	0.25	0.15	0.12	0.18	—	—

* Determined by X-ray fluorescence spectrography.

† Average of 10 microprobe analyses; ‡ average of 14 microprobe analyses.

Table 1, which summarizes bulk analytical and average collophane microprobe analyses for several nodules, supports the above mentioned differences, with Type I nodules having much lower Fe₂O₃ contents than do Type II. Whole rock P₂O₅ contents are low, with a maximum of 15.3% in G7-10, which is slightly below the minimum 18% P₂O₅ content (50% apatite) that is often taken to define a 'phosphorite'¹⁵. Strictly speaking, these are really phosphatic nodules.

Uranium-series age determinations were carried out on several Type I and Type II nodules and one solitary coral (Table 2) by methods previously discussed^{1-4, 16, 17}. Briefly, the basic assumptions of the U-series method relevant to marine phosphorites are: (1) the uranium entered the apatite phase at the time of its formation, and the time of formation was short compared with the age to be measured; (2) the uranium was derived from seawater; (3) the apatite has remained a closed system with respect to the parent and daughter isotopes; and (4) the apatite was initially free of ²³⁰Th. If these constraints are all satisfied, then the ²³⁰Th/²³⁴U ratio provides a reliable estimate

of the time of apatite formation. Previous work^{4, 18} basically validates assumptions (1)–(3), while the small amount of 'initial' ²³⁰Th usually found in marine phosphorites can be corrected for⁴. The results show that all but one of the Type I nodules show finite ages, and three of these are Holocene. If a correction for non-radiogenic ²³⁰Th is made, based on an assumed ²³⁰Th/²³²Th ratio of 4.0 in common thorium⁴, the Holocene nodules would all be <5,000-yr old. Sample G16-3A represents the phosphatic infilling of a solitary coral. The solid base of the coral itself (100% aragonite) was also dated (G16-3B) and we believe this to be the first time that uranium-series dating of coexisting marine phosphorite and corals has been carried out. The results, which are consistent with the depositional sequence as observed under the microscope, coupled with the generally acknowledged reliability of uranium-series dating of corals, considerably strengthens the arguments presented here. Type II nodules are in radioactive equilibrium and thus beyond the limits of the dating method. Sample 1516-A has in fact suffered some leakage of ²³⁴U similar to many nodules from other areas for which a Miocene age has been independently established¹⁷.

There are thus two distinctly different types of phosphorite occurring on the East Australian continental margin. Type II nodules, occurring in depths of <300 m, are relict and are probably of mid-Miocene age, as previously suggested¹³. The Type I nodules, on the other hand, are restricted to a narrow depth interval from 360 to 420 m and seem to have formed by cementation of upper-slope sediments with collophane. The spread of ages displayed by Type I nodules suggests that apatite formation has been a continuous process operating in this area for the last several hundred thousand years, rather than being restricted to high stands of sea level during the late Pleistocene⁴. In fact, one of the nodules (G7-12) has an age of 17,500 yr which is very close to the time of the maximum lowering of sea level associated with the last glaciation. The very young ages obtained for several of the samples indicate that apatite formation within East Australian upper slope sediments has only recently ceased, or is even continuing at present.

We have demonstrated that the East Australian continental margin is an area of Holocene phosphate deposition. The general geological and oceanographical characteristics of the region suggest that this deposit may be a modern analogue of McKelvey's^{5, 6} East Coast phosphorite deposits. Previously, it has proved difficult to explain many East Coast phosphorite deposits in terms of intense coastal upwelling, and other models, such as an estuarine origin have been proposed¹⁹. However, a sizeable phosphate deposit, such as the East Australian continental margin deposit, can be produced, even in an area of only moderate seasonal upwelling. Perhaps the East Australian current, predominantly a shelf feature, has resulted in very low sedimentation rates on the upper slope, thus minimizing 'dilution' of the authigenically formed phosphate.

Table 2 Uranium-series data for phosphatic nodules occurring on the East Australian continental margin

Sample	Type	Depth (m)	U (p.p.m.)	Th (p.p.m.)	²³⁴ U/ ²³⁸ U	²³⁰ Th/ ²³⁴ U	Age (× 10 ³ yr)	Age* _c	²³⁴ U/ ²³⁸ U(IV)	U(IV) (p.p.m.)	U(IV)/U
S12-1	I	376	229	7.3	1.14 ± 0.03	0.075 ± 0.007	≤ 8.5	5 ± 1	—	—	—
S7-3	I	412	153	4.2	1.11 ± 0.01	0.45 ± 0.01	≤ 64	60 ± 2	1.06 ± 0.02	109	0.71
G7-8	I	385	116	3.7	1.12 ± 0.01	0.21 ± 0.01	≤ 25	21 ± 1	1.09 ± 0.01	96	0.83
G7-9	I	385	90	2.7	1.10 ± 0.01	0.33 ± 0.02	≤ 43	40 ± 3	1.07 ± 0.01	66	0.73
G7-10	I	385	132	2.2	1.08 ± 0.01	0.79 ± 0.05	≤ 162	162 ± 23	0.94 ± 0.01	108	0.82
G7-12	I	385	156	3.8	1.10 ± 0.01	0.18 ± 0.01	≤ 21	17.5 ± 1	1.06 ± 0.03	101	0.65
G7-17	I	385	120	4.6	1.02 ± 0.01	0.98 ± 0.04	> 200	> 200	0.81 ± 0.01	94	0.78
G7-29	I	385	87	2.8	1.10 ± 0.01	0.49 ± 0.02	≤ 71	68 ± 2	1.05 ± 0.01	64	0.74
G16-2	I	365	99	9.2	1.17 ± 0.03	0.062 ± 0.006	≤ 7	0	—	—	—
G16-3A	I	365	75	4.7	1.19 ± 0.04	0.079 ± 0.006	≤ 9	2 ± 0.5	—	—	—
G16-3B	Coral	365	4.87	< 0.06	1.14 ± 0.02	0.15 ± 0.007	18	18 ± 0.5	—	—	—
G18-5	II	210	149	3.9	0.99 ± 0.01	0.94 ± 0.028	> 200	> 200	0.72 ± 0.01	117	0.79
1516-A	II	241	87	4.9	0.93 ± 0.01	0.95 ± 0.036	> 200	> 200	—	—	—

Errors quoted based on counting statistics (±1σ).

* Age corrected for common thorium using: $^{230}\text{Th}_c = ^{230}\text{Th}_m - (4 \times ^{232}\text{Th} \exp(-\lambda_{230}t))$, where $^{230}\text{Th}_m$ = measured activity of ²³⁰Th.

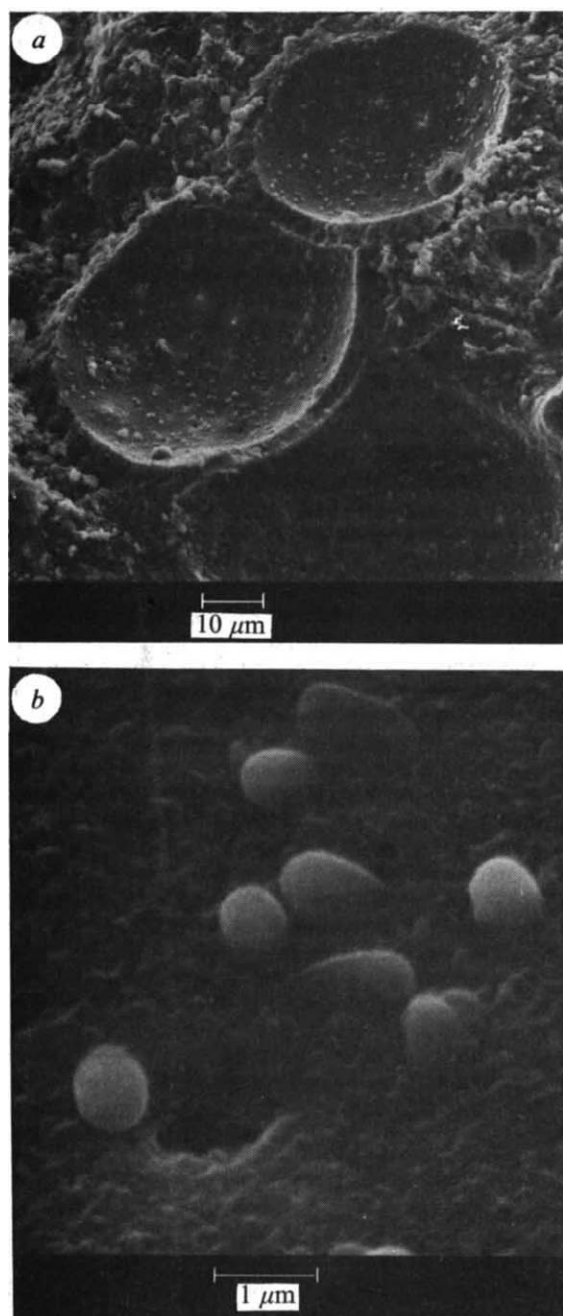


Fig. 2 *a*, Scanning electron micrograph of apatite crystallites on the internal surface of foraminiferal test contained in a phosphatic nodule. (Crystallites confirmed as phosphate-rich by energy dispersive X-ray analysis.) *b*, High magnification of *a* showing relationship of apatite crystals to foraminiferal calcite.

Samples from Locations G7, G16, and G18 were provided by Professor C. C. von der Borch and 1516 by J. Marshall. Others were collected during a CSIRO cruise of RV Sprightly in 1979. Microprobe analyses were performed at Broken Hill Proprietary, Melbourne Research Laboratories and X-ray fluorescence analyses were done by Dr A. R. Milnes. Financial support was provided by the Australian Research Grants Committee. *Note added in proof:* It has been pointed out to us by R. Foster, CSIRO Division of Soils, that the (apatite) crystallites (Fig. 2) could be mineralized bacteria because of the remarkable morphological similarity. If confirmed, this would imply that bacterial activity plays an important role in phosphorite genesis.

Received 24 July; accepted 14 October 1980.

1. Baturin, G. N., Merkulova, K. I. & Chalov, P. I. *Mar. Geol.* **13**, M37-M41 (1972).
2. Veeh, H. H., Burnett, W. C. & Soutar, A. *Science* **181**, 844-845 (1973).

3. Veeh, H. H., Calvert, S. E. & Price, N. B. *Mar. Chem.* **2**, 189-202 (1974).
4. Burnett, W. C. & Veeh, H. H. *Geochim. cosmochim. Acta* **41**, 755-764 (1977).
5. McKelvey, V. E. *U.S. geol. Surv. Bull.* **1252-D**, 1-21 (1967).
6. McKelvey, V. E. & Wang, F. H. *U.S. geol. Surv. Misc. Geol. Inv. Map.* **1-632** (1969).
7. Rochford, D. J. *Div. Fish. Oceanogr. Tech. Pap. No. 33* (CSIRO, 1972).
8. Rochford, D. J. *Aust. J. mar. Freshwat. Res.* **26**, 233-243 (1975).
9. Jones, H. A., Davies, P. J. & Marshall, J. F. *J. geol. Soc. Aust.* **22**, 71-78 (1975).
10. Marshall, J. F. *BMR J. Aust. Geol. Geophys.* **4**, 281-288 (1979).
11. Boland, F. M. & Hamon, B. V. *Deep Sea Res.* **17**, 777-794 (1970).
12. Godfrey, J. S., Cresswell, G. R., Golding, T. J., Pearce, A. F. & Boyd, R. *J. phys. Oceanogr.* **10**, 430-440 (1980).
13. Von der Borch, C. C. *J. geol. Soc. Aust.* **16**, 755-759 (1970).
14. Kress, A. G. & Veeh, H. H. *Mar. Geol.* **36**, 143-157 (1980).
15. Bushinsky, G. I. *Isr. Progr. Sci. Trans. Jerus.* (1969).
16. Clarke, R. S. & Altschuler, A. S. *Geochim. cosmochim. Acta* **13**, 127-142 (1958).
17. Kolodny, Y. & Kaplan, I. R. *Geochim. cosmochim. Acta* **34**, 324 (1970).
18. Burnett, W. C., Veeh, H. H. & Soutar, A. *SEPM Spec. Publ.* (in the press).
19. Pevear, D. R. *Econ. Geol.* **61**, 251-256 (1966).

Boninite at a continental margin

C. P. Wood

New Zealand Geological Survey, PO Box 30368, Lower Hutt, New Zealand

A lava interbedded with Cretaceous sediments deposited on the continental margin of New Zealand has phenocrystic chromite, olivine, low-Ca and high-Ca pyroxenes with quench morphology, normative quartz, and relatively high MgO, Cr and Ni, but very low Ti and Zr; it is shown here to conform to the term boninite as defined by Cameron *et al.*¹. Such rocks are usually found in geologically oceanic situations, as in the Bonin, and Mariana arc-trench systems of the Western Pacific, or in ophiolites that represent uplifted, oceanic crust formed close to oceanic plate margins, particularly spreading centres, though not necessarily at mid-ocean ridges²⁻⁶. The New Zealand boninite closely resembles certain ophiolitic lavas, yet its association is not ophiolitic, and this part of New Zealand never was, nor ever became, an oceanic spreading centre⁷. Coeval basalts are alkalic, not tholeiitic as is usual and the occurrence warns against the use of basalt petrochemistry as an indicator of tectonic environment in the absence of supporting geological evidence.

Albian sediments of the Mangapokia Formation in eastern Wairarapa, in the North Island of New Zealand, accumulated as submarine fan or fan-delta deposits, probably at shelf depths in a fault-controlled basin⁸. They comprise at least 1,500 m of flysch-like sediments, with sporadic lavas and sills, most of which are spilitic, alkalic basalts, some with camptonitic affinities. However, near Nagahape almost 10 m of boninitic lava and

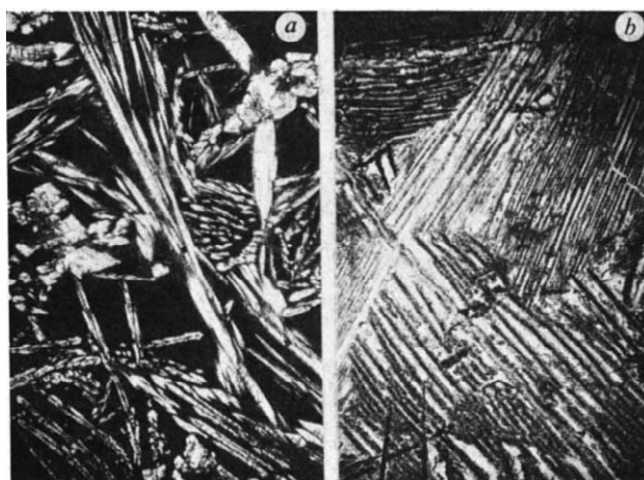


Fig. 1 Textures of Kōpi boninite. *a*, Acicular crystals and skeletal chains of augite, and a few lozenges of pigeonite jacketed by augite. Field width, 1.3 mm. *b*, Spinifex texture; oriented blades of chloritized pigeonite rimmed by augite, and a few, dark, cross-cutting olivine dendrites. Matrix is devitrified glass. Field width, 1.3 mm.

Table 1 Electron probe analyses of representative phenocrysts and average groundmass, and average of six whole-rock analyses (XRF) of Kopi boninite

	1	2	3	4a	4b	5	6	7a	7b
SiO ₂	0.12	55.58	55.89	51.44	51.47	50.20	55.98	52.26 (0.13)	P 900 (100)
TiO ₂	0.08	0.04	0.02	0.18	0.08	0.24	0.16	0.18 (0.01)	Cr 1133 (207)
Al ₂ O ₃	13.37	1.74	1.59	5.51	4.67	7.10	18.80	12.84 (0.24)	Ni 224 (21)
Fe ₂ O ₃	3.85	—	—	—	—	—	—	2.96 (1.08)	Rb 4.8 (1.6)
FeO	18.01	7.11	8.30	7.43	6.73	8.21	7.78	6.11 (0.89)	Sr 39 (4)
MnO	—	0.19	0.18	0.22	0.12	0.16	0.12	0.17 (0.01)	Y 7.7 (1.6)
MgO	10.25	32.29	30.80	18.72	16.83	17.08	4.13	10.97 (0.46)	Zr 9 (2)
CaO	0.04	1.79	2.87	15.55	10.22	16.68	9.75	10.43 (0.30)	Nb <2
Na ₂ O	—	0.04	0.04	0.08	0.10	0.14	3.03	1.43 (0.28)	Ba 63 (30)
K ₂ O	—	—	—	—	0.01	—	0.25	0.12 (0.04)	
NiO	0.07	0.11	0.09	0.04	0.01	0.01	—	—	
Cr ₂ O ₃	53.97	0.87	0.53	0.16	0.36	0.08	—	—	
Total	99.76	99.76	100.31	99.33	99.60	99.90	100.00	97.47	

Column

- 1 Magnesiochromite, Fe₂O₃ calculated assuming stoichiometry.
- 2 Bronzite (En_{86.0}Fs_{10.6}Wo_{3.4}).
- 3 Magnesian pigeonite in phenocryst core (En_{82.1}Fs_{12.4}Wo_{5.5}).
- 4 Augites in phenocryst rims; a, low-Ca end of range (En_{55.0}Fs_{12.2}Wo_{32.8}); b, high-Ca end of range (En_{48.9}Fs_{11.0}Wo_{40.1}).
- 5 Augite rim of spinifex-textured pyroxene.
- 6 Average composition of fine-grained groundmass (nine positions; defocused beam) recalculated to 100%.
- 7 Average boninite analysis, standard deviations in parentheses; a, major elements in wt %, 1.92 (0.61) % H₂O also present; b, trace elements in p.p.m.

hyaloclastite (known as Kopi boninite) occur within a sequence of coarse quartzofeldspathic sediments derived from a continental hinterland. A water depth of 200–700 m at emplacement is suggested by the bulk sulphur content of 940 p.p.m., and average maximum vesicle diameter of 1.5–2 mm (refs 9, 10). Quartz, chlorite, calcite and laumontite in brecciated boninite indicate zeolite facies alteration, though dense rock is little affected.

Pillow lava with conspicuous, random needles of pyroxene predominates, though spinifex-textured rock is also present (Fig. 1). Magnesiochromite crystallized first, and is ubiquitous in trace amounts; with average ratios of Cr/(Cr + Al) = 0.71, and Mg/(Mg + Fe²⁺) = 0.51 it falls at the low-Cr, low-Mg margin of the compositional range of boninite and basaltic komatiite chromites¹. Minor olivine (0.3–1.1 modal %) followed spinel, forming mainly subhedral crystals (<0.5 mm), but with some lantern-and-chain dendrites, particularly in the spinifex rock. Olivine has been replaced completely by quartz, calcite and iron oxides.

There are three types of phenocryst pyroxene, the earliest being rare, unzoned calcic bronzite in the range Mg/(Mg + Fe) = 0.86–0.89 (analysis 2, Table 1). Composite clinopyroxene crystals are the dominant phenocrysts (25.5–38.4 modal %), and in both habit and chemistry they closely resemble the pyroxenes in some rapidly cooled komatiites¹¹. They form hollow needles up to 70 × 1 mm which show simple zoning, with a multiply twinned core of pigeonite rimmed by optically continuous augite. Core compositions are continuous with those of bronzite (Fig. 2), straddling the boundary between calcic clinobronzite (En₈₅Fs₁₁Wo₄) and highly magnesian pigeonite (En₇₆Fs₁₄Wo₁₀). The crystal forms suggest rapid crystallization from a super-cooled liquid to produce these metastable clinopyroxenes, which plot in the miscibility gap between stable high-Ca and low-Ca pyroxenes.

Augite jackets the subcalcic clinopyroxene and sometimes mantles the hollow core; it also occurs as single-phase microphenocrysts and in skeletal chains. In the spinifex rock, oriented blades comprise augite rims on a chloritized core (Fig. 1). Compositions range from En_{55.9}Fs_{13.4}Wo_{30.7} to En_{46.9}Fs_{11.9}Wo_{41.2}, and as the skeletal and spinifex pyroxenes fall within the range of rim compositions (Fig. 2), it seems the composite crystals represent quench overgrowth of calcic clinopyroxene on an earlier, subcalcic core.

Groundmasses comprise fan-spherulitic integrowths mostly too fine to differentiate, but occasionally coarse enough to identify plagioclase (An_{67–79}), clinopyroxene and magnetite

(TiO₂ ≈ 7.0%). Interstitial matrix in the spinifex rock varies from devitrified glass, through axiolitic spherulites to well developed, pin-wheel pyroxene spherulites. The average, bulk-matrix analysis (analysis 6, Table 1) resembles aluminous basaltic andesite. The whole-rock analysis (analysis 7a, Table 1) is less siliceous than boninite from the Bonin Islands, but resembles the basaltic lavas found in some ophiolite complexes such as Troodos², and Betts Cove⁵. Cameron *et al.*¹ class all these as boninite, the name adopted here, but other terms such as basaltic komatiite³, low-Ti basalt⁴, and magnesian quartz tholeiite⁶ have been used for the same rocks. Whatever the name, the combination of relatively high MgO, Cr and Ni, and extremely low TiO₂ and Zr in a quartz-normative

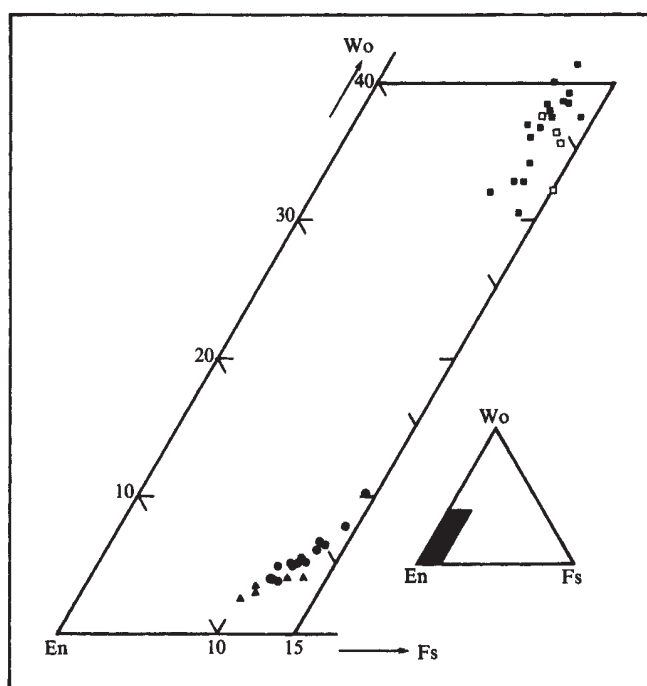


Fig. 2 Compositions of pyroxenes in Kopi boninite plotted in part of the system MgSiO₃–FeSiO₃–CaSiO₃. ▲, Bronzite phenocrysts; ●, cores of composite phenocrysts; ■, rims of composite phenocrysts; □, skeletal and spinifex rim pyroxenes.

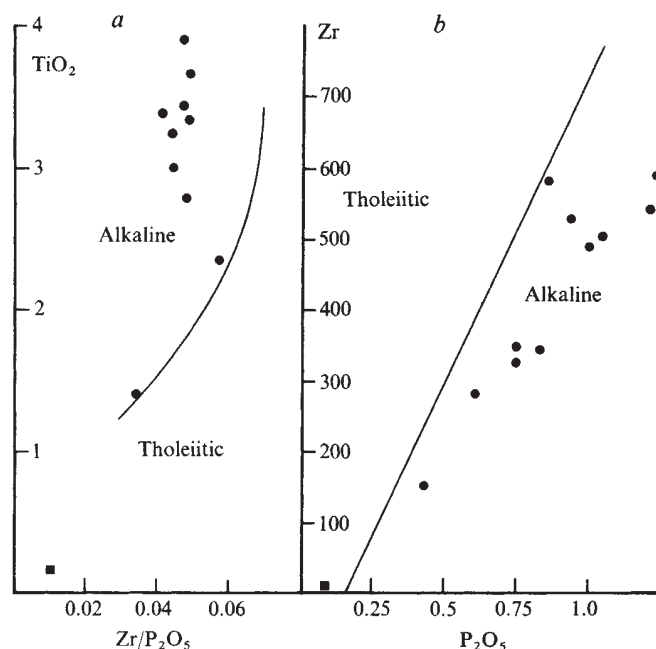


Fig. 3 Kopi boninite (■) and coeval alkaline basaltic rocks (●) plotted using immobile elements to discriminate between alkaline and tholeiitic fields (after ref. 14). *a*, TiO_2 (wt%) against the ratio $\text{Zr}/\text{P}_2\text{O}_5$ (p.p.m.). *b*, Zr (p.p.m.) against P_2O_5 (%).

($Q = 6.0\%$), basic composition places the lava clearly within a spectrum of basaltic to andesitic rocks that are otherwise known only from modern oceanic areas, or ancient ophiolites that were segments of oceanic lithosphere.

The early Cretaceous sediments of eastern Wairarapa are thought to overlie or to represent the youngest part of a sequence of several kilometres of marine sediment of continental provenance, deposited more or less continuously throughout the Mesozoic. An Albian tectonic event deformed the soft sediments and initiated open-shelf deposition⁸ which continued till late Tertiary time, when the present uplift and deformation began. Minor intraformational slumping may have occurred during deposition of Mangapokia sediments, but there is no evidence to suggest allochthonous emplacement of the lavas. Thus there is no indication that the area was anything other than a faulted, continental margin at the time the lava was erupted.

In the discussion of genesis of low-Ti magmas, there is general agreement that multistage melting is necessary to give extreme depletion of incompatible elements^{1,4-6}. Olivine tholeiite and derivatives, produced by voluminous first-stage melting, are the habitual associates of most low-Ti lavas. However, lavas coeval with the Kopi boninite are alkalic basalts and derivatives, with marked enrichment in P, Ti, Y, Zr and Nb (Fig. 3), presumably generated by deep, low-volume, first-stage melting of upper mantle. Eruption of one or more such magma batches would have efficiently depleted a portion of mantle which then could have risen as a small diapir to melt at shallower depths and produce boninite.

If such a process did occur it is perhaps unusual that the association of boninite with alkalic basalt is so rare. The closest comparison seems to be in the Othris Mountains of Greece, where Cameron *et al.*¹ reported boninite in the Agrilia Formation. There, the associated volcanics are mildly undersaturated, alkaline types¹² erupted at a continental margin during a period of slow splitting, possibly caused either by continental rifting, or the initiation of a marginal ocean basin and nearby subduction zone¹³. It is probable that the eastern margin of the early Cretaceous, New Zealand continent was deeply fractured enough to allow boninite magma to erupt, but even if continental splitting had started it clearly failed to develop. It is certain that a spreading centre never formed at the site of deposition,

and though there is a modern subduction zone to the east of northern New Zealand, the existence of one in early Cretaceous times is a moot point.

I thank P. R. Moore, I. G. Speden, and A. J. Tulloch for discussions and advice, and T. R. Grammer and N. Orr for technical help. Most analytical work was carried out in the Geology Department of Victoria University of Wellington.

Received 15 August; accepted 28 October 1980.

1. Cameron, W. E., Nisbet, E. G. & Dietrich, V. J. *Nature* **280**, 550-553 (1979).
2. Smewing, J. D., Simonian, K. O. & Gass, I. G. *Contr. Miner. Petrol.* **51**, 49-64 (1975).
3. Simonian, K. O. & Gass, I. G. *Bull. Am. geol. Soc.* **89**, 1220-1230 (1978).
4. Sun, S.-S. & Nesbitt, R. W. *Geology* **6**, 689-693 (1978).
5. Coish, R. A. & Church, W. R. *Contr. Miner. Petrol.* **70**, 29-39 (1979).
6. Duncan, R. A. & Green, D. H. *Geology* **8**, 22-26 (1980).
7. Stevens, G. R. & Suggate, R. P. in *The Geology of New Zealand*, 727-745 (Government Printer, Wellington, 1978).
8. Moore, P. R. & Speden, I. G. *N.Z. J. Geol. Geophys.* **22**, 417-433 (1979).
9. Moore, J. G. & Schilling, J.-G. *Contr. Miner. Petrol.* **41**, 105-118 (1973).
10. Jones, J. G. *Am. J. Sci.* **267**, 181-195 (1969).
11. Arndt, N. T. *Am. Miner.* **64**, 856-864 (1979).
12. Hynes, A. *Can. J. Earth Sci.* **11**, 842-853 (1974).
13. Smith, A. G. *et al. Ecol. geol. Helv.* **68**, 463-481 (1975).
14. Winchester, J. A. & Floyd, P. A. *Earth planet. Sci. Lett.* **28**, 459-469 (1976).

Coefficients of relatedness in sociobiology

Richard E. Michod

Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, Arizona 85721

W. D. Hamilton

Museum of Zoology, University of Michigan, Ann Arbor, Michigan 48109

A much-discussed, quantitative criterion for the spread of an altruistic gene is Hamilton's rule^{1,2}

$$\frac{c}{b} < R \quad (1)$$

where *c* and *b* are additive decrement and increment to fitness of altruist and recipient, respectively, and *R* is a measure of genetic relatedness between the two individuals. When rearranged as $-c.1 + b.R > 0$, the rule can be interpreted as requiring that the gene-caused action increase the 'inclusive fitness' of the actor. Since its introduction, Hamilton's rule and the attendant concept of inclusive fitness have gained increasing acceptance and use among biologists and have become integral in the field now named sociobiology. However, the essentially heuristic reasoning used in deriving these concepts, along with the lack of a complete specification even in the original outbred model², have led to many investigations into the population genetical underpinnings of Hamilton's rule³⁻¹⁸. On the basis of these considerations, several reports^{3,10,12,13,18} have proposed new formulae for *R*. These formulae have no obvious relation to each other or to the coefficients originally suggested by Hamilton. This proliferation of coefficients is undoubtedly confusing to many and the net effect may be to generate distrust both of the rule and of the notion of inclusive fitness. Our purpose here is to show that these various formulae for *R* (refs 3, 10, 12, 13, 18), although independently derived, are actually the same.

What is *R* in equation (1)? On the basis of a particular outbred model, Hamilton^{1,2} claimed that Wright's¹⁹ 'coefficient of relationship' was the required *R*. However, later, giving a heuristic development but no further model, Hamilton^{20,21} modified his identification of *R* with Wright's coefficient by claiming that equation (1) requires, in principle, a regression coefficient of

genotype of recipient on genotype of altruist, whereas Wright's coefficient is the corresponding correlation coefficient. Such a correlation coefficient will often be the same as the regression coefficient but differs when the interactants are inbred to different extents¹⁸. These discussions^{1,2,20,21} implied that R was independent of gene frequency and selection, and that equation (1) held for inbred populations (with the regression coefficient as R).

However, in recent models³⁻¹⁷, R is taken to be the threshold value of c/b , below which the cost-benefit ratio must be for the gene to increase. The R formulated in this operational way need not necessarily correspond to any simple measure of genetic relationship. Some of these models^{3-11,14,15,17} specify the whole mating process with respect to interactions within families. Other models^{12,13,16} maintain the generality of Hamilton's² original approach, and apply to interactants of arbitrary relationship, but fail to specify the population and mating system processes which give rise to R . If Hamilton's rule is to be useful, it must turn out that R , as formulated by these various approaches, must not vary widely as gene frequency or the parameters of selection change, except as implied by the simple proportionality to c/b . In many cases of interest, R is constant^{4,7-9,15}, but in others^{3,4,7,10-13,16} dependencies on gene frequency and dominance enter into calculations of R .

Our analysis here will not consist of any derivations of the coefficients from the gene frequency dynamics, as that is done from different points of view in the various works to be considered^{3,10,12,13}. Instead, we will simply convert these various coefficients into a common symbolism and show that they are, in fact, identical. However, before doing so, we should acknowledge an important aspect of the derivations: either selection

must be weak so that standard identity coefficients can be used as measures of genetic relationship, or rather special conditions must be imposed on the models so that there is no selection at internal points of the pedigree patterns studied.

The genetic relationship between two diploid individuals, X and Y , at a single locus can be summarized by the following nine 'condensed' identity coefficients $\Delta_1, \Delta_2, \dots, \Delta_9$, where Δ_i is the probability of genetic identity event i obtaining (Fig. 1 upper left-hand corner)²². These coefficients consider identity states among all four alleles present in X and Y . The traditional two-allele coefficients, interpreted as probabilities, are

$$f_X = \Delta_1 + \Delta_2 + \Delta_3 + \Delta_4 \quad f_Y = \Delta_1 + \Delta_2 + \Delta_5 + \Delta_6 \quad (2a)$$

$$f_{XY} = \Delta_1 + \frac{1}{2}(\Delta_3 + \Delta_5 + \Delta_7) + \frac{1}{4}\Delta_8 \quad (2b)$$

where f_X and f_Y are the inbreeding coefficients of X and Y , respectively, and f_{XY} is the 'coefficient of consanguinity' between X and Y , or the probability that a random allele from X is identical by descent with a random allele from Y at the locus of interest. Letting X denote the altruist and Y the recipient, the coefficient which we will use to relate the others may now be defined as

$$R = \frac{f_{XY}\alpha + (\Delta_1 + \frac{1}{2}\Delta_3)(1-\alpha)}{\frac{1}{2}\alpha + f_X(1-\frac{1}{2}\alpha)} \quad (3)$$

where $\alpha = 2q + 2h - 4qh$, p is the frequency of the non-altruist allele A with $q = (1-p)$ the frequency of the altruistic allele a , and h is the probability that a heterozygote performs an altruistic act. If we express the phenotypes as the probabilities of behaving altruistically (0, h , 1 corresponding to AA , Aa , aa),

Fig. 1 Joint and marginal distribution of genetic identity states with genotypes. The nine 'condensed' identity states are given in the upper left-hand corner with probabilities Δ_i ($i = 1, 2, \dots, 9$), $\sum_i \Delta_i = 1$. The alleles of X are on top and the alleles of Y on the bottom. A line connecting alleles indicates identity by descent. Given the occurrence of each identity state, the distribution of the genotypes of X and Y are given within the large side and top three boxes, respectively. For example, if identity state 5 obtains, then X is AA , Aa , aa with probabilities p^2 , $2pq$, q^2 , respectively; the like probabilities for Y are p , 0 , q , respectively. Given the distribution of identity states, the marginal distribution of genotypes is given in the top left corner of each of these six large boxes. The joint distribution of genotypes are given in the nine large boxes towards the bottom right of the figure. For example, if identity state 3 obtains, then

$$\begin{aligned} P_{00} &= p^2, & P_{01} &= pq, \\ P_{03} &= P_{10} = P_{11} = P_{12} = P_{20} = 0, \\ P_{21} &= pq & \text{and} & P_{22} = q^2. \end{aligned}$$

Y's genotype

AA

Aa

aa

X alleles → 
Y alleles → Δ_1



Δ_2 Δ_3 Δ_4 Δ_5
 Δ_6 Δ_7 Δ_8 Δ_9

$$P_{Y0} = p^2 + f_Y pq$$

			p
p	p^2	p^2	p
p	p^2	p^2	p^2

$$P_{Y1} = 2pq(1 - f_Y)$$

			—
—	$2pq$	$2pq$	—
—	$2pq$	$2pq$	$2pq$

$$P_{Y2} = q^2 + f_Y pq$$

			q
q	q^2	q^2	q
q	q^2	q^2	q^2

AA

$$P_{X0} = p^2 + f_X pq$$

			p
p	p	p	p^2
p^2	p^2	p^2	p^2

$$P_{00}$$

			p
p^2	p^2	p^3	p^2
p^3	p^2	p^3	p^4

$$P_{01}$$

			—
—	pq	$2p^2q$	—
—	—	p^2q	$2p^3q$

$$P_{02}$$

			—
pq	—	pq^2	—
p^2q	—	—	p^2q^2

X's genotype

Aa

$$P_{X1} = 2pq(1 - f_X)$$

			—
—	—	—	$2pq$
$2pq$	$2pq$	$2pq$	$2pq$

$$P_{10}$$

			—
—	—	—	pq
$2p^2q$	—	p^2q	$2p^3q$

$$P_{11}$$

			—
—	—	—	—
—	$2pq$	pq	$4p^2q$

$$P_{12}$$

			—
—	—	—	pq
$2pq^2$	—	pq^2	$2pq^3$

aa

$$P_{X2} = q^2 + f_X pq$$

			q
q	q	q	q^2
q^2	q^2	q^2	q^2

$$P_{20}$$

			—
pq	—	p^2q	—
pq^2	—	—	p^2q^2

$$P_{21}$$

			—
—	pq	$2pq^2$	—
—	—	pq^2	$2pq^3$

$$P_{22}$$

			q
q^2	q^2	q^3	q^2
q^3	q^2	q^3	q^4

Table 1 Joint and marginal distributions of X_g and Y_g and X_p and Y_p

		Y			
		Y_g	0	$\frac{1}{2}$	1
		Y_p	0	h	1
X_g	X_p				
0	0		P_{00}	P_{01}	P_{02}
$\frac{1}{2}$	h		P_{10}	P_{11}	P_{12}
1	1		P_{20}	P_{21}	P_{22}
			$P_{Y0} = p^2 + f_Y pq$	$P_{Y1} = 2pq(1 - f_Y)$	$P_{Y2} = q^2 + f_Y pq$
					1

P_{ij} is the joint distribution of genotypes i and j and P_{X_i} , P_{Y_i} are the marginal distributions of the genotypes of X and Y , respectively ($i, j = 0, 1, 2$). From the table, the following expressions can be calculated for use in calculations concerning equations (5): $E(X_p) = hP_{X1} + P_{X2}$, $E(Y_p) = hP_{Y1} + P_{Y2}$, $E(Y_g) = \frac{1}{2}P_{Y1} + P_{Y2} = q$, $E(X_g) = \frac{1}{2}P_{X1} + P_{X2} = q$, $E(Y_g X_p) = \frac{1}{2}hP_{11} + hP_{12} + \frac{1}{2}P_{21} + P_{22}$, $E(X_p X_g) = \frac{1}{2}hP_{X1} + P_{X2}$, $E(X_g Y_g) = \frac{1}{2}P_{11} + \frac{1}{2}P_{12} + \frac{1}{2}P_{21} + P_{22}$, $E(X_g^2) = \frac{1}{4}P_{X1} + P_{X2}$.

and the genotypes as the fractions of altruistic alleles (0, $\frac{1}{2}$, 1 corresponding to AA , Aa , aa), then α is simply the regression of phenotype on genotype in a randomly mating population. This convention regarding the genotypes and phenotypes will be used throughout. With a change in notation, R as given by equation (3) is identical to the coefficient derived by Flesness and Holtzman¹³ (equation (10), where $P(H) = \Delta_1 + \frac{1}{2}\Delta_3$ and their $h = \frac{1}{2}\alpha$). It is important to note that if the gene effects are additive, if there is no inbreeding or if $p = q = 0.5$, then equation (3) reduces to the coefficient proposed by Hamilton^{20,21}:

$$R = \frac{2f_{XY}}{1 + f_X} \quad (4)$$

Consequently, the regression coefficient (equation (4)) will be a good estimate of R in many cases of interest. We now show that the coefficients derived elsewhere in the literature^{3,10,12,18} are identical to equation (3).

Below, we index genotypes with a single subscript i , with $i = 0, 1, 2$ corresponding to genotypes AA , Aa , aa , respectively. Let P_{ij} be the joint distribution of genotypes i and j in the population with i being the genotype of X and j the genotype of Y (Fig. 1, Table 1). Let X_g , X_p and Y_g , Y_p denote the genotype and phenotype of X and Y , respectively.

Orlove³ and Orlove and Wood¹⁰ derive $R = \text{Cov}(X_g, Y_p) / \text{Cov}(X_g, X_p)$ in an outbred model of kin selection in haplodiploid families. Their coefficient^{3,10} equals

$$R = \frac{\text{Cov}(Y_g, X_p)}{\text{Cov}(X_g, X_p)} \quad (5)$$

if there is no inbreeding. However, as we now show, equation (5) is more generally correct, as well as being the most concise and intuitive representation of equations (3) or (17).

To relate equation (5) to equation (3), we must first calculate the required covariances. The information needed to do so is given in Table 1, in which the genotypes and phenotypes of X and Y are given along with the joint and marginal probabilities. Using this information to evaluate the covariances in equation (5), we obtain directly

$$\frac{\text{Cov}(Y_g, X_p)}{\text{Cov}(X_g, X_p)} = \frac{\frac{1}{2}P_{11}h + hP_{12} + \frac{1}{2}P_{21} + P_{22} - (\frac{1}{2}P_{Y1} + P_{Y2})(hP_{X1} + P_{X2})}{\frac{1}{2}hP_{X1} + P_{X2} - (hP_{X1} + P_{X2})(\frac{1}{2}P_{X1} + P_{X2})} \quad (6)$$

From Table 1, we can also obtain for future use

$$\text{Cov}(Y_g, X_g) = \frac{1}{4}P_{11} + \frac{1}{2}P_{12} + \frac{1}{2}P_{21} + P_{22} - (\frac{1}{2}P_{X1} + P_{X2})(\frac{1}{2}P_{Y1} + P_{Y2}) \quad (7)$$

$$\text{Var}(Y_g) = \frac{1}{2}qp(1 + f_Y) \quad (8a)$$

and

$$\text{Var}(X_g) = \frac{1}{2}qp(1 + f_X) \quad (8b)$$

Substituting equation (7) into the numerator of equation (6)

$$\text{Cov}(Y_g, X_p) = \text{Cov}(Y_g, X_g) + (h - \frac{1}{2})(\frac{1}{2}P_{11} + P_{12} - qP_{X1}) \quad (9)$$

The denominator of equation (6) can be simplified directly to

$$\text{Cov}(X_p, X_g) = pq[\frac{1}{2}\alpha + f_X(1 - \frac{1}{2}\alpha)] \quad (10)$$

As shown in Fig. 1 (see also refs 12, 22, 23), it is possible to express the joint probabilities of the interactions, P_{ij} , in terms of the condensed identity coefficients and the gene frequency. In particular,

$$\begin{aligned} P_{11} &= 2pq(\Delta_7 + \frac{1}{2}\Delta_8 + 2pq\Delta_9) \\ P_{12} &= pq(\Delta_5 + 2q\Delta_6 + q\Delta_8 + 2q^2\Delta_9) \end{aligned} \quad (11)$$

These formulae (equation (11) and Fig. 1) are only strictly correct for neutral genes; however, it is hoped that they are approximately correct if selection is weak. Substituting equation (11) into equation (9) and recalling

$$P_{X1} = (1 - f_X)2pq$$

we obtain after some algebra

$$\text{Cov}(Y_g, X_p) = \text{Cov}(Y_g, X_g) + \frac{1}{2}pq(\alpha - 1)[\Delta_5 + \Delta_7 + \frac{1}{2}\Delta_8] \quad (12)$$

Letting r_{XY} be the correlation between the genotypes of X and Y , we have

$$\text{Cov}(Y_g, X_g) = r_{XY}[\text{Var}(X_g) \text{Var}(Y_g)]^{1/2} \quad (13)$$

However¹⁹,

$$r_{XY} = \frac{2f_{XY}}{[(1 + f_X)(1 + f_Y)]^{1/2}} \quad (14)$$

On substituting equations (8) and (14) into equation (13), we obtain

$$\text{Cov}(Y_g, X_g) = f_{XY}pq$$

and so equation (12) becomes

$$\text{Cov}(Y_g, X_p) = f_{XY}pq + \frac{1}{2}pq(\alpha - 1)(\Delta_5 + \Delta_7 + \frac{1}{2}\Delta_8) \quad (15)$$

After rearranging equation (15) and using equation (2b)

$$\text{Cov}(Y_g, X_p) = pq[f_{XY}\alpha + (1 - \alpha)(\Delta_1 + \frac{1}{2}\Delta_3)] \quad (16)$$

On substituting equations (16) and (10) into equation (5), we obtain equation (3) as was to be shown.

Michod¹² (equation (3)) derived the following coefficient for use in Hamilton's rule

$$R = \frac{2xpp + (1 - x)(1 - 2q)\rho'}{1 + x - 2q} \quad (17)$$

$$\rho = \frac{(\Delta_1 + \frac{1}{2}\Delta_3)}{q + pf_X} + \frac{q(\Delta_5 + \Delta_7 + \frac{1}{2}\Delta_8)}{q + pf_X} \quad (18a)$$

and

$$\rho' = \frac{\Delta_5 + \Delta_7 + \frac{1}{2}\Delta_8}{1 - f_X} \quad (18b)$$

which are Jacquard's conditional genic structures^{12,22,23}. These conditional coefficients in equation (18) give the probability with which X can predict, on the basis of pedigree ties, the allelic distribution in gametes of Y . This probability depends on

whether X is homozygous (equation (18a)) or heterozygous (equation (18b)). In addition, the variable x appearing in equation (17) is the frequency of altruistic homozygotes among the total class of altruists and is

$$x = \frac{q^2(1-f_x) + f_x q}{q^2(1-f_x) + q f_x + (1-f_x)2pqh} \quad (19)$$

Equation (19) corrects a mistake in Table 1 of Michod¹² where it was thought that the average inbreeding coefficient (averaged over the frequencies of X and Y in the population) should be used in equation (19). This is incorrect due to the conditional nature of the altruism. As the gene is only expressed in X , when X is in a certain relationship to Y (the relationship being summarized by the Δ values), it is only the frequency of homozygous altruists among all X individuals which matter in equation (19). This fact alone accounts for the incorrect impression, given in Fig. 2b of that paper¹², that dominant genes could never increase in frequency. (Fig. 2a of ref. 12 is unaffected by these considerations.) After substituting equations (19) and (18) into equation (17) and collecting the Δ values in terms of f_{xy} , equation (3) is obtained after some rearrangement.

In conclusion, the various coefficients derived in the literature^{3,10,12,13,18} for use as R in Hamilton's rule are equivalent. These coefficients were originally derived from very different points of view. Given the results here, these various approaches support each other and suggest that the one pleomorphic coefficient [equations (3), (5) or (17)] is, indeed, correct.

This work was supported in part by NSF grant DEB 79-10191 to R.E.M. We thank Bob Abugov for discussion and comment.

Received 14 April; accepted 16 October 1980.

1. Hamilton, W. D. *Am. Nat.* **97**, 354-356 (1963).
2. Hamilton, W. D. *J. theor. Biol.* **7**, 1-51 (1964).
3. Orlove, M. J. *J. theor. Biol.* **49**, 289-310 (1975).
4. Levitt, P. R. *Proc. natn. Acad. Sci. U.S.A.* **72**, 4531-4535 (1975).
5. Scudo, F. M. & Ghiselin, M. T. *J. Genet.* **62**, 1-31 (1975).
6. Charnov, E. L. *J. theor. Biol.* **66**, 541-550 (1977).
7. Cavalli-Sforza, L. L. & Feldman, M. W. *Theor. Population Biol.* **14**, 268-281 (1978).
8. Charlesworth, B. *J. theor. Biol.* **72**, 297-319 (1978).
9. Wade, M. J. *Proc. natn. Acad. Sci. U.S.A.* **75**, 6154-6158 (1978); *Am. Nat.* **113**, 399-417 (1979).
10. Orlove, M. J. & Wood, C. L. *J. theor. Biol.* **73**, 679-686 (1978).
11. Craig, R. *Evolution* **33**, 319-334 (1979).
12. Michod, R. E. *J. theor. Biol.* **81**, 223-233 (1979).
13. Flesness, N. R. & Holtzman, R. C. *Am. Nat.* (in the press).
14. Michod, R. E. *Genetics* (in the press).
15. Michod, R. E. & Abugov, R. *Science* **210**, 667-669 (1980).
16. Charlesworth, B. in *Dahlem Workshop on Social Behaviour: Hypothesis and Empirical Tests* (ed. Markl, H.) 11-26 (Verlag Chemie, Weinheim, 1980).
17. Abugov, R. & Michod, R. E. *J. theor. Biol.* (in the press).
18. Michod, R. E. & Anderson, W. W. *Am. Nat.* **114**, 637-647 (1979).
19. Wright, S. *Am. Nat.* **56**, 330-338 (1922).
20. Hamilton, W. D. in *Man and Beast: Comparative Social Behaviour* (eds Eisenberg & Dillon) 57-91 (Smithsonian Institution Press, Washington, DC, 1971).
21. Hamilton, W. D. *A. Rev. Ecol. Syst.* **3**, 193-232 (1972).
22. Jacquard, A. *The Genetic Structure of Populations* (Springer, New York, 1974).
23. Elston, R. C. & Lange, K. *Ann. Hum. Genet.* **39**, 493-496 (1976).

Proline and valine—cues which stimulate grasshopper herbivory during drought stress?

Brent M. Haglund

Montana Department of Natural Resources and Conservation, Biology Department, Montana State University, Bozeman, Montana 59717

Insect herbivores, including members of Orthoptera and Lepidoptera, benefit from increased nitrogen in their food, particularly if it is in the form of easily digested amino acids^{1,2}. I report here experimental evidence that grasshoppers detect and preferentially feed on grasses treated with the amino acids proline and valine, which commonly increase in plants under drought stress³. This ability may lead to insect concentrations on drought-stressed and nitrogen-enriched plants and thus exacerbate acridid population outbreaks through enhanced growth and survival.

Ecological studies of grasshoppers and locusts have often considered weather factors to be important determinants of population size or growth rate⁴⁻⁹. Dry soil conditions in particular seem to favour population increases of many economically important Orthopteran species¹⁰. It has been hypothesized that grasshopper growth rates, and consequent survival, increase sharply in drier conditions when perennial grasses have elevated nitrogen (N) levels in parts above ground¹. Insect herbivores are generally likely to benefit from increased dietary N (ref. 2).

Such elevated nitrogen levels may be the result of increased concentrations of proline³. A highly soluble amino acid, proline does not hamper the activities of many photosynthetic enzymes and is able to provide osmotic adjustment in conditions of drought stress or salt accumulation¹¹⁻¹⁷. Glycine and valine may function similarly¹⁸. Leaf laminae accumulate the largest quantities of the newly synthesized proline although it is also found in most other above-ground parts¹⁹. Genetic varieties of plants which accumulate larger concentrations of free proline tend to survive water stress better and grow faster after stress is relieved²⁰.

Mobile herbivores which can detect and concentrate their feeding on N-enriched plants will be favoured relative to conspecifics lacking these abilities. These abilities will be particularly important in environments with food patches of varying N quality. For example, in the northern Great Plains of North America, soil moisture is usually limiting to plant growth in summer. Here forage quality—normally influenced by rainfall—is patchy because thundershowers are local and infrequent.

I conducted a field experiment in a native Montana grassland to test the hypothesis that chewing insects can detect plants with elevated amino acid nitrogen levels within the range of levels that occur naturally. A completely randomized two-factor design was replicated three times. Plots of 2 m² separated by 4 m buffers were treated. Treatment levels by factor were Factor A (irrigation): none and 25 mm soil surface irrigation; Factor B (foliar sprays): none, 1.6 litre distilled H₂O, Pro + Val in equal amounts, mixed amino acids in equal amounts (Ala, Arg, Gly, His, Lys, Met, Phe), Gly and gelatin. For the last four treatment levels, amino acids or protein were dissolved in 1.6 litre distilled H₂O with a concentration equal to 0.7534 g N (except Gly which equalled 7.534 g N). The treatment level was chosen to lie within the range of plant water stress responses and raise the estimated N standing crop by 10%, based on 84 g m⁻² standing crop plant biomass at an estimated 2% total N.

The irrigation treatments were carried out on 9 July 1979 between 0700 and 1000 h; foliar spraying was done by hand on 10 July 1979 from 1230 to 1700 h. Except for the Gly treatment, which was 10 times more concentrated, the maximum possible individual amino acid concentration was 0.93% of plant dry matter compared with measured proline levels of 1% or greater for barley¹⁴, Bermuda grass¹⁵, tobacco¹⁶, clover²¹ and several halophyte species^{11,17} under water stress. Proline can accumulate to as much as 10-20% of shoot dry weight in *Triglochin maritima*¹⁷. I estimated that less than one quarter of the spray from any of the treatments contacted the leaves due to low plant density at the site.

The dependent variable measured was percentage of leaves chewed on stems of *Agropyron smithii* with four intact leaves, a C₃ grass, and *Bouteloua gracilis* with five intact leaves, a C₄ grass, stratified by leaf position enumerated from youngest to oldest. Virtually complete collections of stems of these two species were gathered by hand on 6 September 1979. The total number of stems sampled for each of the foliar spray treatments is shown in Fig. 1. Only chewed leaves were counted—broken leaf tips were not recorded.

Chewing was affected by two amino acid treatments but was unaffected by irrigation or the remaining amino acid sprays (χ^2 test, $P < 0.001$). Because irrigation treatments showed no influence on chewing ($P > 0.3$) data were pooled across water levels in the subsequent analysis. Paired t -test analysis of percentage of leaves chewed by leaf position indicated

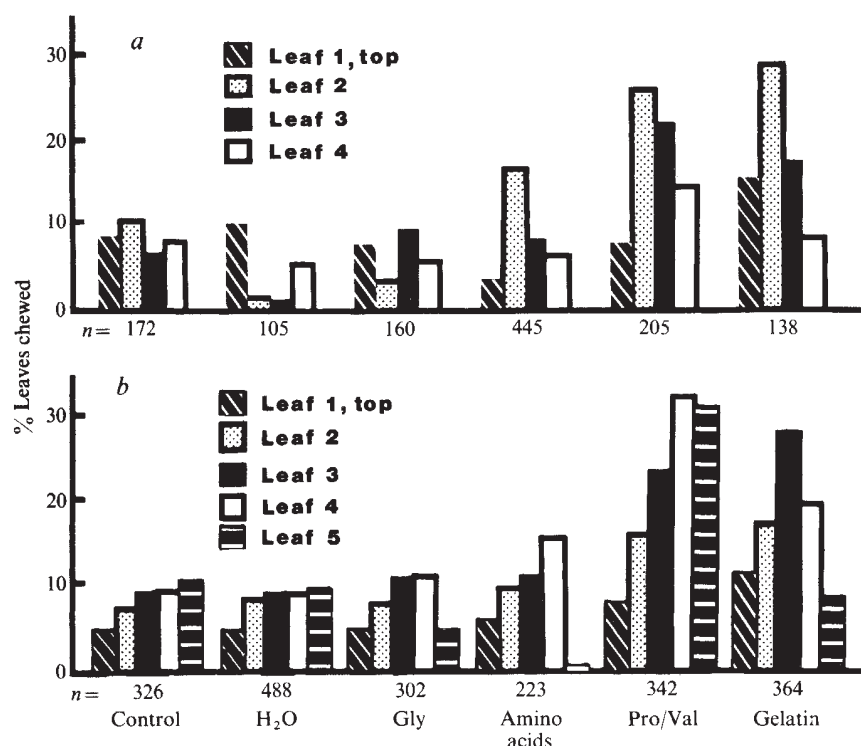


Fig. 1 Percentage of leaves chewed by grasshoppers in response to amino acid foliar spray treatments on two grass species: *Agropyron smithii* (a) and *Bouteloua gracilis* (b). Each bar represents the percentage of leaves at each position that were chewed. Leaf position is assigned from the top to the bottom of the stem. Sample size (n) is the total number of stems assayed within each treatment.

significant differences ($P < 0.01$) on the second and third leaves of *A. smithii* and on the third, fourth and fifth leaves of *B. gracilis* with Pro and Val. With the gelatin treatment, only the second leaf of *A. smithii* and the third and fourth leaves of *B. gracilis* showed greater numbers of chews. At the time of spraying the first leaf of both species had not emerged; the fourth leaf of *A. smithii* was already brown and probably dead.

Grasshoppers were not responding to an improvement in plant condition due to N enrichment. Forty randomly selected plants from each foliar treatment were measured for stem length, leaf length (numbers two and three) and colour (completely brown or some green remaining). With analysis of variance no difference in these variables could be detected among the treatments ($P > 0.2$).

I conclude that grasshoppers (over 95% of the chewing insects on the plots in July and August and dominated by *Opeia obsura*, *Trachyrachis kiowa* and *Melanoplus sanguinipes*) can detect either proline or valine on plant tissue. They selected grass treated with proline and valine and also grass treated with gelatin containing proline (15%) and valine (3%) residues. Although glycine is an important component (45%) of gelatin, glycine alone had no effect on insect chewing. The possibility that grasshoppers are capable of chemodetection is supported by the documentation²² of the presence of amino acid receptors in other chewing insects, the Lepidoptera. Aphids are also known to detect amino acids²³. Furthermore, proline is reported to be a feeding attractant to the snail *Biomphalaria glabrata*²⁴.

Proline is used by animals in various ways including intermediate energy metabolism, enzyme synthesis and, in large amounts, for connective tissue. It can also be used for osmoregulation. Insects able to find and consume proline in above average quantities are likely to grow faster. I suggest that proline accumulations will affect chewing insects much less in wet years or in conditions of augmented soil moisture through weather modification or irrigation.

Management implications of this plant-herbivore interaction include the possibility of lowered insect herbivory to the extent that weather modification or irrigation could relieve plant

drought stress. Plants intimately coevolved with their insect herbivores may find partial relief from drought stress by herbivory because it reduces leaf area, allowing water to become more concentrated in remaining leaf tissue^{25,26}. However, crop plants genetically engineered for greater proline production in the hopes of raising their drought tolerance and thereby increasing their yield may be at increased risk to insect pests²⁷.

I thank T. W. Weaver III for ideas. This research was funded by the High Plains Cooperative Program, Montana Department of National Resources and Conservation grant no. 14-06-D-7577, Water and Power Resources Service, US Department of Interior. Land was made available by the Livestock and Range Research Station, Agricultural Research Service, US Department of Agriculture.

Received 17 April; accepted 15 October 1980.

- White, T. R. C. *Oecologia* **22**, 119–134 (1976).
- Slansky, F. Jr & Feeny, P. *Ecol. Monogr.* **47**, 209–228 (1977).
- Hsiao, T. C. A. *Rev. Pl. Physiol.* **24**, 519–570 (1973).
- Andrewartha, H. G. & Birch, L. C. *The Distribution and Abundance of Animals* (University of Chicago Press, 1954).
- Betts, E. *Anti-Locust Mem.* **6**, 1–25 (1961).
- Dempster, J. P. *Biol. Rev.* **38**, 490–529 (1963).
- Edwards, R. L. *Can. Ent.* **92**, 619–624 (1960).
- Rodell, C. F. *Ecology* **58**, 227–245 (1977).
- Uvarov, B. P. *Trans. ent. Soc. Lond.* **79**, 1–247 (1931).
- Gage, S. H. & Mukerji, M. K. *Envir. Ent.* **6**, 469–479 (1977).
- Cavaliere, A. J. & Huang, A. H. C. *Am. J. Bot.* **66**, 307–312 (1979).
- Hanson, A. D. & Tully, R. E. *Planta* **145**, 45–52 (1979).
- Singh, T. N. *et al. Aust. J. Biol. Sci.* **26**, 45–56 (1973a).
- Singh, T. N. *et al. Nature new Biol.* **236**, 188–190 (1972).
- Barnett, N. M. & Naylor, A. W. *Pl. Physiol.* **41**, 1222–1230 (1966).
- Mizusaki, S. *et al. Archs Biochem. Biophys.* **105**, 599 (1964).
- Stewart, G. R. & Lee, J. A. *Planta* **120**, 279–289 (1974).
- Naylor, A. W. in *Water Deficits and Plant Growth* (ed. Kozlowski, T. T.) 241–254 (Academic, New York, 1972).
- Singh, T. N. *et al. Aust. J. Biol. Sci.* **26**, 57–63 (1973b).
- Singh, T. N. *et al. Aust. J. Biol. Sci.* **26**, 64–72 (1973c).
- Routley, D. G. *Crop Sci.* **6**, 358–361 (1966).
- Schoonhoven, L. M. *Nature* **221**, 1268 (1969).
- Van Emden, H. F. *Insect/Plant Relationships* (Symp. R. ent. Soc., London, 1972).
- Uhazy L. S. *et al. Science* **201**, 924–926 (1978).
- McNaughton, S. J. *Am. Nat.* **113**, 691–703 (1979).
- Wisniol, K. *J. Range Mgmt* **32**, 194–195 (1979).
- Marx, J. L. *Science* **206**, 1168–1169 (1979).

Insect populations respond to fluctuating environments

David A. Jillson

Department of Zoology, University of Vermont, Burlington, Vermont 05405

Models of populations inhabiting regularly fluctuating environments suggest that the nature of the population fluctuation is determined by the relative sizes of the characteristic response time of the population, T_R , and the period of the environmental variation. For example, May's¹ model of a density-regulated population inhabiting a sinusoidally varying environment showed that when the characteristic response time was long relative to the period of environmental fluctuation, Y , the population averaged out environmental fluctuations, whereas for T_R short relative to Y , the population tended to track environmental variations. Characteristic response time is a measure of how quickly a system returns to equilibrium following a disturbance, and for the logistic model and its immediate relatives, T_R equals the inverse of population growth rate, r . The present study investigated the responses of populations of the flour beetle, *Tribolium castaneum*, cultured in a series of regularly fluctuating environments. Responses to environmental fluctuations were striking at all periods of environmental fluctuations, but population density declined as environmental period lengthened.

Environmental fluctuations were created by altering the amount of culture medium (95% unbleached wheat flour, 5% brewer's yeast by weight) between 32 and 8 g at regular intervals. This experimental regime represents fluctuations in carrying capacity, for *Tribolium* population size is known to be positively correlated with the amount of culture medium². *T. castaneum* populations were cultured in each of five experimental regimes—a constant 20-g environment, and fluctuating environments with 4-, 8-, 12- and 16-week periods. The periods were chosen to represent intervals ranging from the equivalent of, to much longer than an estimate of *T. castaneum* characteristic response time. *Tribolium* characteristic response time was calculated by taking the inverse of population growth rate. This procedure is accurate if the logistic model is an adequate representation of *Tribolium* population dynamics and was used as simple approximation to T_R . Estimates of *T. castaneum* population growth rates derived from age-specific survivorship and fecundity schedules range from 0.7 to 1.3 per week³⁻⁶, corresponding to T_R of approximately 1 week. Estimates of r taken from exponential growth phases of *T. castaneum* population growth data⁶⁻⁸ yield longer response times of 2-6 weeks, and may be more indicative of attainable rates of growth.

Thirty cultures of the corn oil-sensitive strain of *T. castaneum*⁹⁻¹⁶ were initiated with 30 adults in a 1:1 sex ratio and 75 unsexed small larvae. Populations were censused at 2-week intervals: larvae, pupae and adults were counted and returned to fresh medium of the appropriate amount. All populations were cultured in 20-g medium for the initial 18 weeks of the experiment to allow the populations to approach demographic and numerical equilibrium. At week 18, six populations were randomly assigned to each of the five treatments, and the series of environmental fluctuations began, continuing until week 64.

Regular though small fluctuations were characteristic of the mean population size over time in the constant 20-g environment (Fig. 1). These intrinsic fluctuations were partly caused by cannibalistic interactions¹⁵⁻¹⁸. A significant linear regression of mean population size with time showed that populations increased modestly at a rate of only four animals per week.

In contrast, populations in the 4-week treatment increased in density substantially. During weeks 18 to 54, populations increased on average by 21 animals per week. Fluctuations were also seen in this treatment—local increases in population size occurred during periods of 32-g medium, and, except for the decline from week 54 to 56, every decrease occurred in 8-g medium. Bursts of recruitment occurred in the relatively abundant 32-g medium, enhancing the number of small larvae successfully recruited. After two weeks of 8-g medium, 32 g were

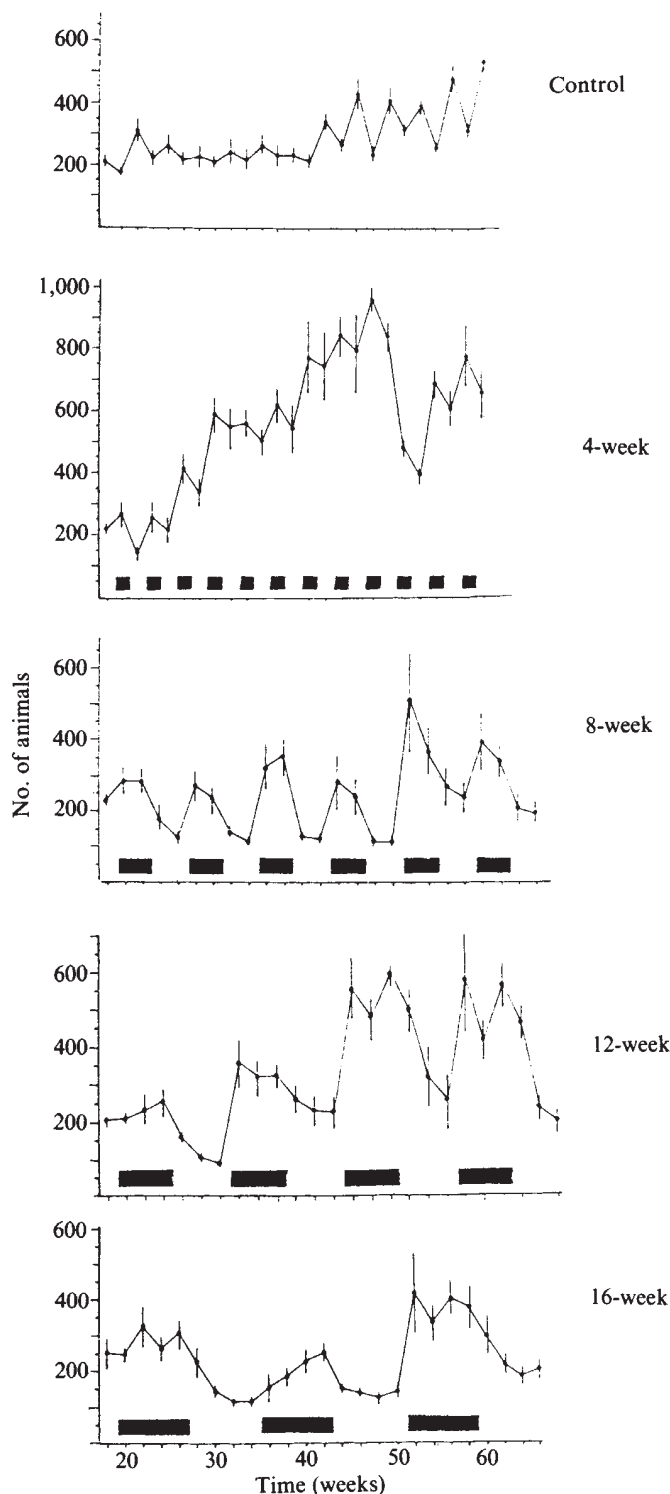


Fig. 1 Mean number ($\pm$ s.e.) of animals in constant and fluctuating environments. Shaded bars in fluctuating treatments represent periods of 32-g medium, intervening periods had 8-g medium.

available again, the small larval cohort had become large larvae and were able to pupate in abundant medium where cannibalism was reduced. It is most interesting that the 4-week period of environmental fluctuation was ideally suited to the 4-week *T. castaneum* developmental period by maximizing the recruitment of small larvae and minimizing cannibalism on the vulnerable pupae. This treatment attained the highest population densities of the experiment, approximately twice that of the other treatments. This does not agree with predictions that fluctuations in carrying capacity reduce the long-term average population number below that attained in a constant environment^{1,19,20}. Wool and Sverdlov²¹ reported a similar discrepancy—*Tribolium* productivity tended to be higher in temporally variable than in constant environments. Predictions of a reduction in mean population size were premised on the logistic model, which fails to incorporate the effects of cannibalism, an important factor in *Tribolium* population dynamics^{17,18}.

Populations inhabiting the 8-week treatment responded to environmental fluctuations differently from those in the 4-week cycle. Periods of abundant resources supported high densities and vice versa. In contrast to the 4-week treatment, these populations increased only moderately with time. Although bursts of recruitment occurred in 32-g conditions, each cohort of small larvae suffered high losses due to cannibalism when pupation occurred in crowded 8-g conditions.

In the 12-week treatment, mean population size again responded to environmental fluctuations, with population abundance directly proportional to the amount of culture medium. Population peaks tended to contain local dips, reflecting intrinsic patterns of cannibalism on eggs by a cohort of large larvae. Populations in the 16-week treatment also clearly mimicked the physical regime. The three complete cycles of culture medium resulted in three population cycles. Populations in the 12- and 16-week treatments were able to recruit young and undergo pupation in a sequence of 32-g medium, but the next cycle of recruitment was delayed by 6 and 8 weeks respectively by the intervening 8-g period, effectively reducing mean population size.

The proportion of adults in the control and 4-week cycle was relatively constant, and did not display any strong correlation with environmental variability. In contrast, the age distribution in the 8-, 12- and 16-week treatments followed environmental fluctuations, with 80–90% adults found during periods of sparse culture medium, alternating with 20–30% adults in periods of 32-g medium. These demographic oscillations were predominantly caused by recruitment of substantial numbers of immatures in 32-g medium whereas adult numbers remained relatively constant over time.

Tribolium populations clearly exhibited changes in density and age distribution in concert with temporal changes in resource availability. In addition, the large mean population size attained in the 4-week treatment was caused by a particular temporal pattern of environmental variation which suited this animal ideally. Although the mean amount of culture medium with time was the same in all treatments, 20 g, differing sequences of resource distribution yielded large differences in population size.

This work was supported by a University of Vermont Institutional Award.

Received 9 July; accepted 13 October 1980.

- May, R. M. in *Theoretical Ecology* (ed. May, R. M.) 4–25 (Saunders, Philadelphia, 1976).
- Chapman, R. N. *Ecology* **9**, 111–122 (1928).
- Leslie, P. H. & Park, T. *Ecology* **30**, 469–477 (1949).
- Mertz, D. M. *Ecol. Monogr.* **39**, 1–31 (1969).
- Young, A. M. *Ecology* **51**, 602–619 (1970).
- Jillson, D. A. thesis, Univ. Rhode Island (1977).
- Park, T., Leslie, P. H. & Mertz, D. B. *Physiol. Zool.* **37**, 97–162 (1964).
- Park, T. *Physiol. Zool.* **27**, 177–238 (1954).
- Costantino, R. F., Bell, A. E. & Rogler, J. C. *Nature* **210**, 221–222 (1966).
- Costantino, R. F., Bell, A. E. & Rogler, J. C. *Heredity* **22**, 529–539 (1967).
- Costantino, R. F., Rogler, J. C. & Bell, A. E. *Heredity* **23**, 477–483 (1968).

- Costantino, R. F., Mumma, R. O. & Bruszewski, T. E. *Heredity* **25**, 411–418 (1970).
- Costantino, R. F. & Rowe, P. M. *Heredity* **29**, 247–252 (1972).
- Scully, A. M. M. & Costantino, R. F. *Can. J. Genet. Cytol.* **17**, 585–590 (1975).
- Moffa, A. M. & Costantino, R. F. *Genetics* **87**, 785–805 (1977).
- Jillson, D. A. & Costantino, R. F. *Am. Nat.* **116**, 206–219 (1980).
- Lloyd, M. *Ecology* **49**, 245–259 (1968).
- Desharnais, R. A. & Costantino, R. F. *Can. J. Genet. Cytol.* (in the press).
- Nisbet, R. M. & Gurney, W. S. C. *J. theor. Biol.* **56**, 459–475 (1976).
- May, R. M. *Stability and Complexity in Model Ecosystems* (Princeton University Press, New Jersey, 1973).
- Wool, D. & Sverdlov, E. *Evolution* **30**, 119–129 (1976).

Multiple extraocular photoreceptive areas on genitalia of butterfly *Papilio xuthus*

Kentaro Arikawa, Eisuke Eguchi*, Akihiro Yoshida & Kiyoshi Aoki

Life Science Institute, Sophia University, 7-1 Kioicho, Chiyoda-ku, Tokyo, Japan

* Department of Biology, Yokohama City University, Kanazawa-ku, Yokohama, Japan

Extraocular photoreception has been reported in several orders of insects, and direct photoreception by the central nervous system (CNS) has been shown to be related to the circadian clock in many cases¹. For example, CNS photoreceptors have been found in the brain and the last abdominal ganglion of the cockroach, *Periplaneta americana*, and in the brain of the silkworm, *Hyalophora cecropia*². In other cases dermal light sensitivity has been deduced from behavioural responses. For instance, the larvae of *Tenebrio molitor* avoid light even after decapitation³. The antennae of *Aphis fabae* seem to be the site of light sensitivity responsible for the animal's photokinetic activity⁴. Dermal light sensitivity is demonstrable in all the tergites of the larvae of *Acilius japonicus* and *Dytiscus marginalis*, particularly well developed in the region of terminal abdominal spiracles⁵. We report here, however, the electrophysiological response and morphological characteristics of presumptive photoreceptive sites on the genitalia of the Chinese yellow swallowtail butterfly, *Papilio xuthus* L.

Electrophysiological recordings were made on summer forms of *P. xuthus* collected from a wild population in June 1980. For anatomical studies we used spring forms reared on a semi-artificial diet in the laboratory at 20 °C and with a 8L:16D light regime.

We found that two (N5, N6) of the six pairs (N1–N6, numbered from the most anterior pair) of nerve roots belonging to the last abdominal ganglion contained afferent photosensitive axons. Recordings could be made easily from cut proximal ends of these nerves by means of suction electrodes (Fig. 1a): responses consisted of sustained trains of spikes.

In white light, impulse frequency is related to log *I* (light intensity) by an S-shaped curve. The latent period varies inversely with *I* (Fig. 1a). Spectral sensitivity was determined for the photosensitive units by using 17 monochromatic light bands of 380–700 nm; it peaked at about 400 nm (Fig. 1b). Only one class of units has been recognized.

However, exploration with a 1-mm spot of light demonstrated two distinct photoreceptive areas on each side of the genital region, one (P1) sending afferents to root N6, and the other (P2) connecting with the CNS through N5. P1 and P2 are located in different areas specific to males and females. By anatomical isolation (for P1) and cutting of other nerves (for P2) the photoreceptor sites can be localized as follows.

In males ♂ P1 lies in the scaphium whereas ♂ P2 is on the fulcrum inferior (Fig. 2a). The scaphium is a sclerotized plate-like structure on the dorsal surface of the tuba analis, and the fulcrum inferior is a ventral part of the terminal diaphragm of

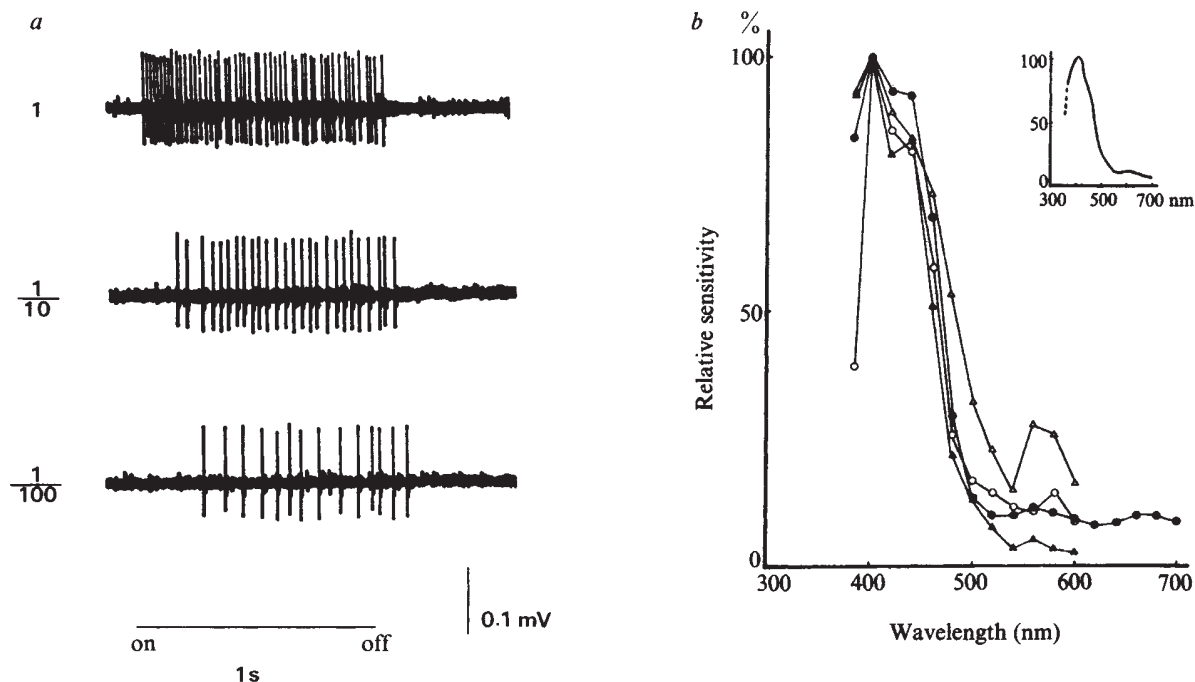


Fig. 1 Electrophysiological responses. *a*, Typical spike train evoked by a 1-s flash of white light. Numbers at the left are the relative light intensities ($I \sim 5,000$ lx). Recording from N5 of the last abdominal ganglion of a male *P. xuthus*. This nerve receives afferents from the fultura inferior (P2). Similar responses were obtained from N5 as well as N6 in both sexes. *b*, Spectral sensitivity of four individuals: $\circ$ and $\bullet$, males recorded from N6; Δ , female recorded from N6; $\triangle$, female recorded from N5. Spike numbers at $\lambda_{\max} = 400$ nm were taken as 100%. Inset, averaged curve.

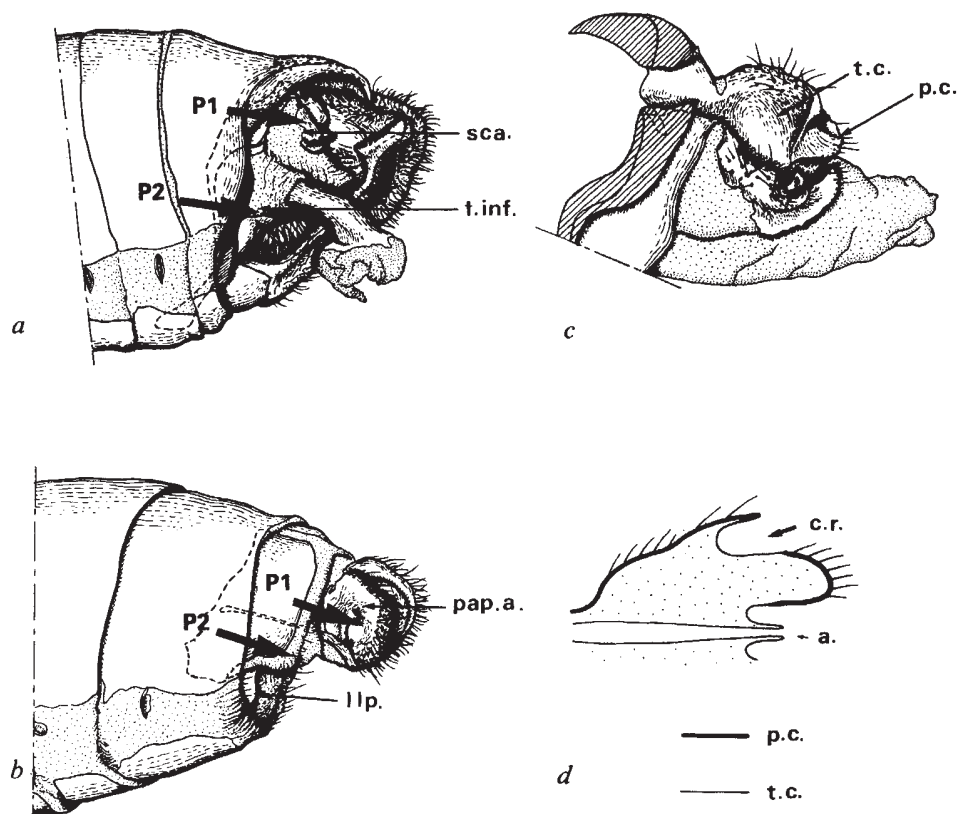


Fig. 2 Location (four arrows in *a* and *c*) and external features of *P. xuthus* photoreceptive sites identified by response to stimulation by a 1-mm spot of light. *a*, Tip of male abdomen seen laterally from the left side with left valva removed; *b*, tip of female abdomen seen laterally from the left side; *c*, detail of male scaphium; *d*, schematic horizontal section of the concave light sensitive area of the papilla analis. sca., Scaphium; f.inf., fultura inferior; pap.a., papilla analis; llp., lamella postvaginalis; p.c., pigmented cuticle; t.c., transparent cuticle; c.r., concave region; a., anus.



Fig. 3 Presumed photoreceptive cell in the male scaphium. *a*, Overall view (scale = 1 μ m); *b*, detail of lamellated structure in the cytoplasm of the cell (scale = 0.1 μ m). Specimens were prefixed for 4–5 h with 2.0% glutaraldehyde and 2.0% paraformaldehyde in 0.1 M cacodylate buffer solution at pH 7.4 at room temperature. After washing, the tissue was post-fixed for 2–3 h in 2.0% OsO_4 in the same buffer, dehydrated in an acetone series and embedded in Epon. Cut with a diamond knife, sections were stained with uranyl acetate and lead citrate and observed with a JEM 100B electron microscope. N, nucleus; B, basement membrane; L, lamellated body; E, epidermal cell; O, oil droplet.

male butterflies. In females ♀ P1 lies on the papilla analis and ♀ P2 lies laterally on the membrane between the papilla analis and the lamella postvaginalis (Fig. 2*b*). The papillae analis covered with chitinous hairs are strongly sclerotized lobes between which the anus and ostium oviductus open to the exterior.

Externally these various photoreceptor sites are marked by small areas of transparent cuticle, which elsewhere is pigmented and opaque. Thus the scaphium is pale brown over most of its surface but has a transparent area on its inner margin. Also the papilla analis is strongly pigmented except for a concave posterior region which is transparent and free of hairs (Fig. 2*d*).

When the transparent P1 genital areas are covered with opaque material the afferent responses in N6 are blocked completely. When this opaque material is removed, light-induced spikes reappear as in the normal preparation (Fig. 1). This shows that the photoreceptive elements of P1 are illuminated through its local clear cuticular area. Similar transparent spots as described above, characterize the P2 area but the effect of screening them has not been tested yet.

To localize further the receptor elements involved in ♂ P1 and ♀ P1 we examined these areas by transmission electron microscopy. That revealed, directly under the basement membrane of the epidermis, cells containing well-developed lamellated

organelles (Fig. 3*a*). These are essentially the same under the corresponding transparent cuticle of both ♂ and ♀ P1.

These lamellated bodies occupy a large volume of cytoplasm and may be associated with the photosensitivity. Lamellated bodies resembling those identified in the butterfly scaphium and papilla analis have been observed in the eyes of other invertebrates—for example, in the distal photoreceptor cells of the eye of the scallop *Pecten*⁶ and the proximal region of cephalopod visual cells⁷.

The visual pigment of both vertebrates and invertebrates is embedded in elaborate disk-like or microvillus arrays of photoreceptor membrane. However, the epidermal cells of these butterfly photoreceptors show no such surface cytoplasmic membrane system, suggesting that they are not the light sensitive units in this case. However, there are other cells directly beneath the transparent cuticle and epidermis, with extensive cytoplasmic membrane systems, and they may be the photoreceptive organelles. A more convincing demonstration of such a function would depend on both proof of the appropriate axonal or synaptic connection to supply the observed spikes in N6, and a more detailed study of the lamellated membrane systems. Fine structural changes due to light and dark adaptation, as well as evidence for localization of visual pigment molecules might also prove diagnostic in this case. Whether and how these receptors are functional in copulation or other reproductive behaviour remain to be discovered.

We also obtained responses to white light stimulation of the genitalia with 6 other species of Papilionidae and 10 species from 6 of the 7 other butterfly families found in Japan (Libytheidae have not been tested yet). However, no such responses could be detected electrophysiologically in either nocturnal (two species tested) or diurnal (four species tested) moths.

We thank Mr Jun Emoto, Nanzan University, for critical comments and Dr Talbot H. Waterman, Yale University, and Dr Jean M. Michalec, Sophia University, for reviewing the manuscript. This work was supported in part by a Grant-in-Aid for Scientific Research 411201 and 510906 from the Ministry of Education, Science and Culture of Japan.

Received 7 August; accepted 30 October 1980.

1. Bennett, M. F. in *Handbook of Sensory Physiology* Vol. 7, 6A (ed. Autrum, H.) 641–664 (Springer, New York, 1979).
2. Saunders, D. S. *Insect Clocks* (Pergamon, Oxford, 1976).
3. Tuculesco, J. *Bull. Biol.* no. 4, 1–35 (1933).
4. Booth, C. O. *Nature* **197**, 265–266 (1963).
5. Schöne, H. Z. *vergl. Physiol.* **33**, 63–98 (1951).
6. Barber, V. C. *et al.* *Z. Zellforsch.* **76**, 295–359 (1967).
7. Yamamoto, T. *et al.* *J. Cell Biol.* **25**, 345–359 (1965).

Reversible activation–inactivation of renin in human plasma

Brenda J. Leckie & Norah K. McGhee

Medical Research Council Blood Pressure Unit, Western Infirmary, Glasgow G11 6NT, UK

Renin, an aspartate protease, cleaves the α -globulin angiotensinogen to produce the decapeptide angiotensin I, which is then converted to the vasoactive hormone angiotensin II by the action of a peptidase ‘converting enzyme’. An inactive form of renin^{1–3} sometimes termed prorenin⁴ is present in normal human plasma. Its enzymatic activity is increased by exposure to a pH of 3.0 or 3.3 followed by dialysis towards neutral pH. Only a small proportion of the inactive renin is activated during the acid stage of dialysis, most of the activation apparently taking place during the subsequent dialysis to pH 5.7 (ref. 4) or 7.5 (ref. 5). Furthermore, if inhibitors of serine proteases are added to the plasma, the amount of inactive renin activated by this

dialysis procedure is reduced⁴⁻⁶. These results suggest that acid-activation is mediated by serine proteases. The role of enzymes such as plasma kallikrein, plasmin and renal kallikrein as physiological activators of inactive renin has recently been discussed⁵⁻⁹. In our study of the activation of plasma inactive renin we have now found that, contrary to previous reports^{4,5}, complete activation of inactive renin takes place during the acid stage of dialysis. This activation can be reversed if plasma is rapidly adjusted to pH 7.4 and warmed. The next step in the acid-activation procedure, that is, dialysis to neutral pH, renders the initial acid-activation irreversible. These results were completely unexpected, and we offer an explanation that reassesses the nature of inactive renin and the activation process.

In the first series of experiments we studied the change in the renin concentration of normal male plasma during acid-activation in the presence and absence of soya bean trypsin inhibitor (Fig. 1). Samples of plasma (0.5 ml) were dialysed to pH 3.0 for 24 h at 4 °C (ref. 1) and then taken to pH 7.4 at 4 °C. At intervals, plasma was removed from the dialysis bags and transferred to 50 μ l soya bean trypsin inhibitor (SBTI) solution⁶ to

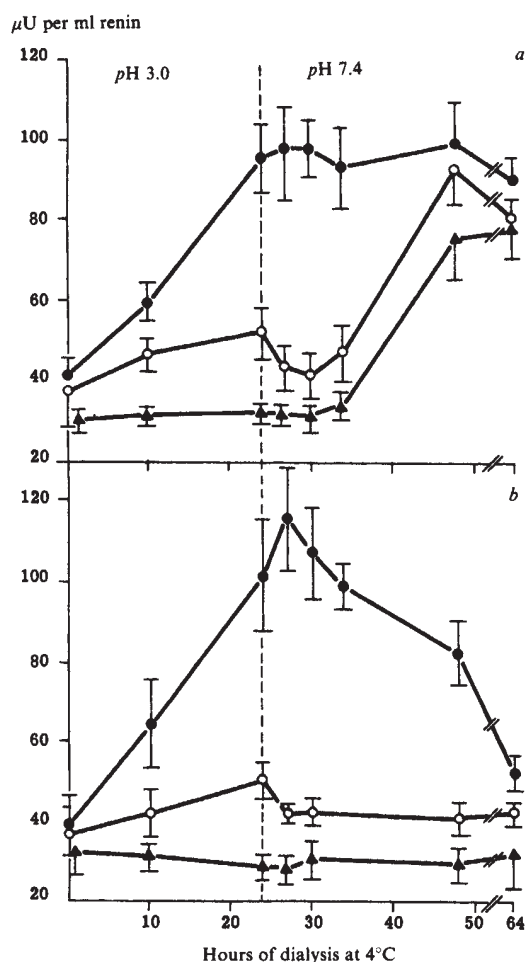


Fig. 1 The renin concentration (ordinate) in normal human plasma during dialysis at 4 °C to pH 3.0 followed by dialysis to pH 7.4. *a*, 0.5-ml samples of plasma were dialysed against 15 ml of pH 3.0 glycine/HCl buffer¹ at 4 °C. At 0, 10 and 24 h, samples were removed from the dialysis bags and added to 50 μ l of a 5 mg ml⁻¹ solution of SBTI to give a final concentration of 19 μ mol per 1 SBTI. 50 μ l of 3 M Tris-HCl buffer was added and the renin concentration assayed immediately by measuring the rate of angiotensin I generation from ox angiotensinogen¹⁰. ●, Immediate assay; ○, assay after 1 h at 37 °C; ▲, assay after 5 h at 37 °C. *b*, As for *a* except that 19 μ mol per 1 SBTI was added to the plasma before the start of dialysis. Mean $\pm$ s.e.m. of four experiments.

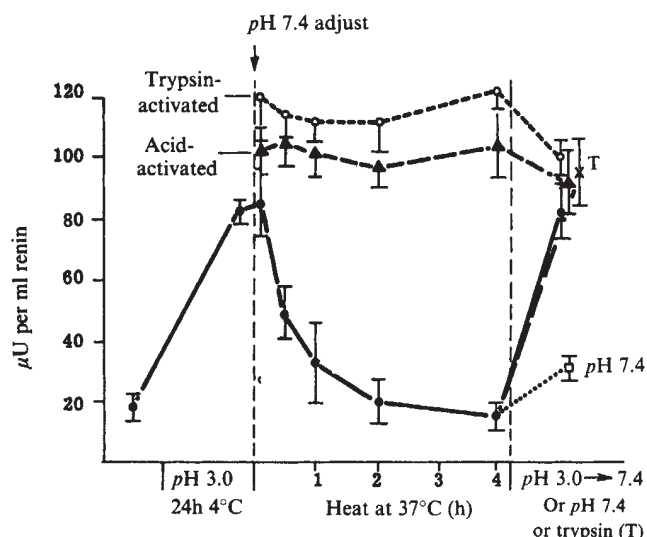


Fig. 2 ●, Plasma was dialysed to pH 3.0 for 24 h at 4 °C then 1/10 volume of 3 M Tris-HCl buffer added and the plasma warmed at 37 °C. After 4 h the remaining plasma was dialysed to pH 3.0 and dialysed back to pH 7.4 at 4 °C. Controls were dialysed to pH 7.4 for 48 h at 4 °C (□). ▲, Plasma was dialysed to pH 3.0 for 24 h at 4 °C and dialysed back to pH 7.4 at 4 °C before the addition of 3 M Tris-HCl buffer and warming. ○, Plasma was trypsin-activated before being treated as described above. Mean $\pm$ s.e.m. of three experiments.

give a final concentration of 19 μ mol l⁻¹ SBTI. The inhibitor was used to prevent any activation of inactive renin by plasma proteases after removal of the sample from the dialysis bag⁶. The pH of the plasma was adjusted to between 7.4 and 7.5 at 4 °C with the addition of 50 μ l 3 M Tris-HCl buffer, pH 7.6, and the samples were immediately assayed for renin¹⁰. The renin was re-assayed after the samples had been incubated for 1 h at 37 °C and again after 5 h. In a parallel series of experiments, 19 μ mol l⁻¹ SBTI was added to the plasma before the start of the dialysis. Samples were dialysed as before, except that no further SBTI was added before adjustment of the pH and the samples were assayed for renin. Figure 1*a* shows the results of experiments in which plasma was dialysed without SBTI and assayed for renin immediately after adjustment to pH 7.4. During dialysis at pH 3.0 the renin concentration increased from 41 $\pm$ 4 μ U ml⁻¹ to 95 $\pm$ 8 μ U ml⁻¹ (mean $\pm$ s.e.m., P < 0.01). During the subsequent dialysis to pH 7.4, the renin concentration did not change: 98 $\pm$ 9 μ U ml⁻¹ at 24 h (that is, 48 h from the start of the experiment) and 89 $\pm$ 6 μ U ml⁻¹ after 40 h (64 h from the start). Thus dialysis to pH 3.0 completely activated the inactive renin in this plasma. We confirmed that activation was complete by treating a separate portion of the plasma with trypsin¹⁰, which gave a renin concentration of 100 $\pm$ 12 μ U ml⁻¹ similar to that of 95 $\pm$ 8 μ U ml⁻¹ obtained with acid alone.

A different pattern was obtained if the samples were allowed to stand at 37 °C for 1 h before being assayed. In this case, the renin concentration apparently increased to only 52 $\pm$ 6 μ U ml⁻¹ after the pH 3.0 stage of dialysis, but to 91 $\pm$ 7 μ U ml⁻¹ after dialysis to pH 7.4. If the plasma was assayed after 5 h at 37 °C there was an apparent lack of activation during the pH 3.0 stage of dialysis. This pattern of activation is similar to that obtained elsewhere^{4,5}.

Figure 1*b* shows the results of the experiments in which SBTI was added to the plasma before the start of dialysis. The activation obtained during the pH 3.0 stage of dialysis was similar to that for renin in plasma dialysed without SBTI. However, during the dialysis to pH 7.4 in the presence of SBTI the renin concentration in the plasma fell slowly over 40 h so that the final value of 58 $\pm$ 5 μ U ml⁻¹ was lower than that after the dialysis to pH 3.0 alone (P < 0.05) and also lower than the

renin concentration in plasma dialysed at pH 3.0 and then taken to pH 7.4 for 40 h in the absence of SBTI ($P < 0.01$). When samples that had been dialysed in the presence of SBTI were warmed at 37 °C and reassayed, no apparent activation of renin was noted throughout both periods of dialysis.

These results showed that the activation of inactive renin induced by acid dialysis could be reversed if the plasma was rapidly neutralized and warmed. We investigated the reversal in a second series of experiments using another sample of pooled normal plasma. The plasma was dialysed to pH 3.0 for 24 h at 4 °C, assayed for renin then adjusted to pH 7.4 with 3 M Tris-HCl buffer at 4 °C and reassayed. The plasma was warmed at 37 °C and the renin concentration measured at intervals. After 4 h at 37 °C the remaining plasma was again dialysed to pH 3.0 for 24 h and dialysed back to pH 7.4 for a further 24 h at 4 °C to determine whether the renin could be re-activated. A control dialysis to pH 7.4 for 48 h at 4 °C was carried out, and the effect of activation by trypsin was also tested. Figure 2 shows that exposure to pH 3.0 increased plasma renin concentration from 19 ± 3 to 82 ± 2 $\mu\text{U ml}^{-1}$. During the incubation at 37 °C the renin concentration fell to the level in untreated plasma. If this incubated plasma was either acid-activated or trypsin-activated the renin concentration was restored to 81 ± 4 $\mu\text{U ml}^{-1}$, or 97 ± 11 $\mu\text{U ml}^{-1}$ respectively. Therefore, the loss of renin on warming at 37 °C is not an irreversible denaturation similar to that of active renin exposed to pH 11.4 for 2 h (ref. 11). Incubated plasma dialysed to pH 7.4 showed a renin concentration of only 36 ± 4 $\mu\text{U ml}^{-1}$. This was higher than the renin levels in untreated plasma but it is possible that cryoactivation⁴ had occurred during the 48 h at 4 °C.

Experiments were also carried out with plasma that had been dialysed to pH 3.0 and dialysed back to pH 7.4 for 48 h at 4 °C

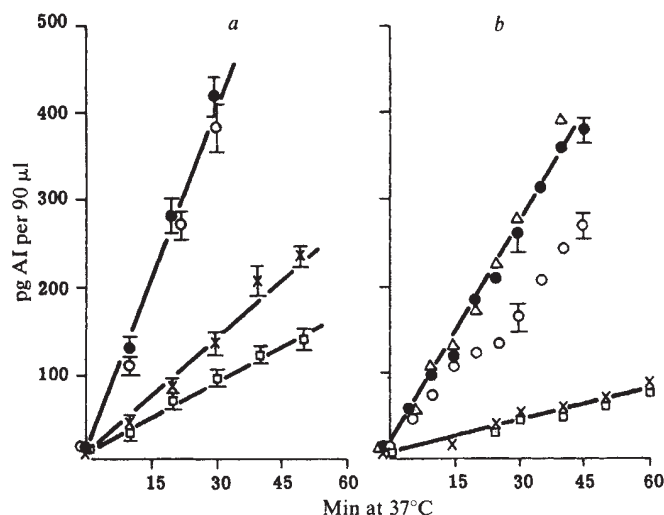


Fig. 3 Rate of production of angiotensin I from ox angiotensinogen by renin in two samples of normal plasma, using an antibody-trapping method¹⁰. Plasma was treated as follows: ●, dialysed to pH 3.0 at 4 °C and dialysed back to pH 7.4 at 4 °C; ○, dialysed to pH 3.0 at 4 °C, rapidly adjusted to pH 7.4 at 4 °C and assayed; □, as for ○ but the plasma was incubated for 5 h at 37 °C before assay; ×, untreated plasma; △, trypsin-activated plasma (sample b only). The pg angiotensin I per 90 µl incubation mixture (ordinate) is plotted against time of incubation (abscissa). The range of the radioimmunoassay is 30–500 pg angiotensin I per 90 µl. Note that all samples of plasma including (○) and (□) show undetectable amounts of angiotensin at the start of the assay and that rate of production of angiotensin I is linear with time except for plasma b treated as for (○) (see text). In plasma a, heating at 37 °C for 5 h reduced the renin concentration below that of untreated plasma although for four separate experiments this was not significant ($P > 0.05$). Bars show s.e. for triplicate assays.

before addition of 3 M Tris-HCl buffer and incubation at 37 °C. This plasma showed no reversal of the high renin concentration when warmed, nor did trypsin-treated plasma (Fig. 2).

The reversible activation-inactivation of renin in plasma was not due to an assay artefact. Falsely high renin concentrations can be obtained if proteolytic enzymes present in the radioimmunoassay mixture destroy the iodinated angiotensin I (results in preparation). False high results could also be obtained if a substance that cross-reacted with our angiotensin I antibody was generated during the dialysis of plasma to pH 3.0. A falsely low result could be obtained if the angiotensin I produced during the renin assay was destroyed by peptidases. However, none of the plasma samples contained detectable amounts of angiotensin I-like material before the start of the incubation with ox angiotensinogen. Thus the high renin concentrations obtained after dialysis of the plasma to pH 3.0 were not due to high 'blank' concentrations of angiotensin I (Fig. 3). The rate of production of angiotensin I was proportional to the time of incubation with ox angiotensinogen (Fig. 3), for all samples of treated plasma, except three out of six samples that had been dialysed to pH 3.0 without subsequent dialysis to pH 7.4. In these, the rate of angiotensin I generation decreased with increasing time of incubation in the renin assay. This cannot be due to an artefact such as substrate exhaustion which should equally affect other samples of the same renin concentration, but might be expected if the 'activated' renin in acid-treated plasma was being inactivated during the renin assay. We measured the recovery of angiotensin I incubated with plasma in the presence of angiotensin I antibody¹⁰. Recovery was >96% during the period of the renin assay, after correction for any generation of angiotensin I in the plasma. Therefore the antibody could protect angiotensin I against peptidases in the treated plasma samples. We considered the possibility of artefacts due to the presence of a protease such as cathepsin D which could generate angiotensin I from ox angiotensinogen. However, cathepsin D is not likely to cleave ox angiotensinogen at our assay pH of 6.9–7.0 (ref. 12). We conclude therefore that our results reflect a genuine change in the enzymatic activity of plasma inactive renin.

Since this study we have heard that another group using a different renin assay method also finds that plasma inactive renin is activated by exposure to acid alone (W. Hsueh, personal communication). Hsueh *et al.*¹³ suggest that acidification alters the structure of the inactive renin molecule in such a way as to render it more susceptible to the action of proteases such as renal kallikrein⁹. We suggest an alternative explanation. Human plasma contains many protease inhibitors. There are high molecular weight forms of renin in mouse plasma that may consist of active renin bound to a plasma protein¹⁴. Furthermore, inactive forms of renin in extracts of pig¹⁵, rabbit¹⁶ and dog¹⁷ kidney consist of active renin bound to a renin 'inhibitor'. The inactive renin in human plasma may be a similar material. Exposure to acid may result in a conformational change in either renin or inhibitor leading to dissociation of the complex. Restoration of alkaline pH and incubation at 37 °C could result in re-association. By contrast, dialysis of the acidified plasma towards pH 7.4 at 4 °C could result in loss of the binding capacity of the inhibitor if plasma serine proteases either destroy the inhibitor by proteolysis or bind to it in competition with renin.

Received 1 September; accepted 27 October 1980.

1. Leckie, B. J. *et al.* *Circulation Res.* **40**, 46–51 (1977).
2. Derckx, F. H. M. *et al.* *Lancet* **ii**, 496–498 (1978).
3. Bord, G. W. *Lancet* **i**, 215–218 (1977).
4. Atlas, S. A., Sealey, J. E. & Laragh, J. H. *Circulation Res.* **43**, Suppl 1, 128–135 (1978).
5. Derckx, F. H. M., Bouma, B. N., Schalekamp, M. P. A. & Schalekamp, M. A. D. H. *Nature* **280**, 315–316 (1979).
6. Leckie, B. J. *Clin. Sci. molec. Med.* **55**, 133s–134s (1978).
7. Sealey, J. E., Atlas, S. A., Laragh, J. H., Silverberg, M. & Kaplan, A. P. *Proc. natn. Acad. Sci. U.S.A.* **76**, 5914–5918 (1979).
8. Rumpf, K. W. *et al.* *Nature* **283**, 482 (1980).
9. Sealey, J. E., Atlas, S. A., Laragh, J. H., Oza, N. B. & Ryan, J. W. *Nature* **275**, 144–145 (1978).
10. Millar, J. A., Leckie, B. J., Morton, J. J., Jordan, J. & Tree, M. *Clinica chim. Acta* **101**, 5–15 (1980).

11. Arakawa, K., Ikeda, M., Fukuyama, J. & Sakai, T. *J. clin. Endocr. Metab.* **42**, 599–602 (1976).
12. Hackenthal, E., Hackenthal, R. & Hilgenfeldt, U. *Biochim. biophys. Acta* **522**, 574–588 (1978).
13. Hsueh, W. A., Carlson, E. J., O'Connor, D. & Warren, S. *Clin. Res.* **28**, 331 (1980).
14. Poulsen, K. *et al. Clin. Sci. molec. Med.* **55**, 125s–128s (1978).
15. Boyd, G. W. *Circulation Res.* **35**, 426–438 (1974).
16. Leckie, B. J. & McConnell, A. *Circulation Res.* **36**, 513–519 (1975).
17. Kawamura, M., Ikemoto, F., Funakawa, S. & Yamamoto, K. *Clin. Sci. molec. Med.* **57**, 345–350 (1979).

Ability of kallikrein to generate angiotensin II-like pressor substance and a proposed 'kinin-tensin enzyme system'

Kikuo Arakawa & Haruo Maruta

Department of Internal Medicine, Fukuoka University School of Medicine, Fukuoka, Japan 814

Pig pancreatic kallikrein liberates kallidin from kininogen¹, whereas trypsin releases bradykinin². Recently, both kallikrein and trypsin have been reported to convert inactive plasma renin to active renin^{3,4}. However, we found that at pH 6.0, trypsin generated an angiotensin II-like pressor substance from human plasma protein in the absence of converting enzyme⁵. This has been isolated and found to have the same amino acid composition as angiotensin II⁶. Thus, *in vitro* trypsin can directly liberate both the depressor, bradykinin, in weak alkaline conditions, and the pressor, angiotensin II, at weakly acidic pH, from the appropriate substrates. We have now investigated whether kallikrein—a serine protease like trypsin—also generates a pressor substance at weakly acidic pH. Our results demonstrate that it does. We therefore suggest that kallikrein may be involved in a direct link between the pressor and depressor systems and we propose the term 'kinin-tensin system' for this sort of one-enzyme system capable of generating both depressor and pressor substances.

Human plasma protein Cohn's fraction IV-4 was used as substrate because it is rich in angiotensinogen⁷, as well as kallidinogen⁸. However, it was not a pure preparation, and was shown to contain renin which can be inactivated by alkali treatment⁹. Possible contamination with prorenin, cathepsin D and angiotensin I-converting enzyme was excluded or eliminated as follows: first, to activate any prorenin present, fraction IV-4 was treated by acid according to Yokosawa *et al.*³, and second, inactivation of the activated renin was achieved by treatment with alkali according to Arakawa *et al.*⁹. Trypsin- or kallikrein-activatable prorenin^{3,4} was similarly treated. As shown in Table 1, the absence of angiotensin I formation indicated that these treatments were effective in removing any activatable prorenin and renin. Acid protease, or cathepsin D activity was not detected using denatured bovine haemoglobin as the substrate¹⁰. Angiotensin I-converting enzyme activity was also not detected when assayed by the method of Cushman and Cheung¹¹. Thus, the alkali-treated substrate was shown to be free from any interfering components.

For pure pig pancreatic kallikrein two components, kallikrein A and B, were revealed by cellulose acetate electrophoresis at pH 8.6 (150 V, 1.5 h). Trypsin and chymotrypsin migrate different and could not be detected. The preparation did not hydrolyse benzoyl-L-tyrosine ethyl ester¹², which indicated that there was no contamination by chymotrypsin, and it was also shown to be free from cathepsin D and angiotensin I-converting enzyme, by the same methods used for the above substrate.

The pretreated human plasma protein fraction IV-4 and the pure kallikrein preparation were incubated at pH 3.0–10.0, and after incubation vasoactive substances (Table 2) were analysed using the rat blood pressure bioassay. The bioassay showed formation of a pressor substance at pH 4.0–6.0, and depressor

Table 1 Proof of non-contamination by trypsin- or kallikrein-activatable prorenin in substrate, human plasma protein fraction IV-4

Total amount of the pressor substance formed from 10 mg of substrate						
Incubation of trypsin-activated fraction IV-4 with renin-free substrate				Incubation of kallikrein-activated fraction IV-4 with renin-free substrate		
Incubation (min)	Bioassay (equiv. to AII ng)	AI-RI (ng)	AII-RI (ng)	Bioassay (equiv. to AII ng)	AI-RI (ng)	AII-RI (ng)
0	0	0	0.04	13.6	0	12.5
15	0	0	0.04	15.2	0	17.5
30	0	0	0.03	13.8	0	16.0
60	0	0	0.03	8.9	0	15.7
120	0	0	0.03	8.0	0	14.0

Before incubation, to activate acid-activatable prorenin which might be contained in fraction IV-4, it was treated by dialysis against 0.05 M glycine-HCl buffer containing 0.1 M NaCl, pH 3.3, for 20 h, followed by a second dialysis against 0.1 M Tris-HCl buffer containing 0.1 M NaCl, pH 7.5, for 2 h. To inactivate the resultant activated renin, it was treated at pH 11.3 with 1.0 M NaOH for 2 h. pH was adjusted to 7.0 with 1.0 M HCl and it was dialysed against distilled water, and lyophilized. All these experiments were done at 4 °C. To activate trypsin- or kallikrein-activatable prorenin, 10 mg of the above-treated fraction IV-4 of 920 ng angiotensin I equivalent were dissolved in 1 ml of 0.1 M Tris-HCl buffer, pH 7.5, and added with 0.4 mg of trypsin (Miles, specific activity 3,000 NF units) or 0.5 mg of hog pancreatic kallikrein (Sanwa Chemicals, specific activity 768 µg equivalents bradykinin per min per mg, pH 8.0). Incubation was at 25 °C for 10 min in the case of trypsin and 1 h for kallikrein. The reaction was stopped by the addition of 5 mg aprotinin (Bayer, 5,000 KIU per mg protein). Finally, 100 µl of the above trypsin- or kallikrein-treated mixture were incubated with 10 mg of the renin-free substrate, or alkali-treated fraction IV-4⁹. The incubation was done in 2 ml of 0.1 M sodium phosphate buffer, pH 6.0, containing 200 µl of 0.1 M Na₂EDTA, 20 µl of 0.45 M 8-hydroxyquinoline sulphate and 5 µl of 10% 2,3-dimercapto-1-propanol as angiotensinase inhibitors, at 37 °C for 0, 15, 30, 60 and 120 min. Reaction was stopped by boiling for 10 min. Products were bioassayed in rat blood pressure, and radioimmunoassayed for angiotensin I (AI-RI) and angiotensin II (AII-RI). Control experiments using 10 mg of acid- and alkali-treated fraction IV-4 without treatment by trypsin or kallikrein showed no formation of pressor substance. Trypsin- or kallikrein-treated fraction IV-4 produced a small amount of angiotensin II, but not angiotensin I.

substance(s) at the other pHs. Pepstatin, a renin inhibitor¹³, did not inhibit the pressor formation (Table 2), indicating that a renin-like enzyme was not involved. In control experiments without kallikrein, the alkali-treated fraction IV-4 did not generate any pressor substance in identical conditions, as reported elsewhere^{5,9}. Kallikrein alone without substrate also did not generate pressor. These data are therefore consistent with the liberation of a vasoconstrictor by kallikrein and parallel previous results with trypsin^{5,6}.

The pressor effect was not only similar in pattern to that of angiotensin II, but was also inhibited by the angiotensin II-antagonist, [1-Sar,8-Ile]-angiotensin II (Fig. 1), but not by SQ 14225, the converting-enzyme inhibitor (Fig. 2). Bioassay value of the pressor activity coincided with the radioimmunoassay value for angiotensin II, but not for angiotensin I, as shown in Table 2. It therefore seems likely that the pressor effect is due to the release of angiotensin II, in contrast to angiotensin I formed from the non-alkali-treated substrate in the absence of kallikrein⁹. Unless unknown mechanisms are involved, it is probably derived from angiotensinogen by a single cleavage. The methods used in the purification of the kallikrein and substrate indicate that the activation of prorenin^{3,4} is not involved. Involvement of tonin is also unlikely because tonin has never been found in any animal organ other than rat submaxillary gland and pure pig pancreatic kallikrein is unlikely to contain it.

Thus both trypsin and kallikrein have been shown to be involved in the direct release of both pressor and depressor

Table 2 Comparison of values of the pressor substance formed at pH values 3–10 by kallikrein assayed by different methods

pH	Total amount of pressor substance formed after 18 h incubation					
	Without pepstatin			With pepstatin		
	Bioassay (equiv. to AII ng)	AI-RI (ng)	AII-RI (ng)	Bioassay (equiv. to AII ng)	AI-RI (ng)	AII-RI (ng)
3	145.0	21	251.0	0	0	16.8
4	1,760.0	10	1,770.0	1,636.0	0	1,664.0
5	1,520.0	8	1,892.0	1,912.0	0	2,580.8
6	370.0	28	340.0	87.2	0	70.0
7	5.5	13	15.7	0	0	14.8
8	75.0	0	15.4	0	0	4.9
9	54.0	0	6.6	0	0	2.7
10	69.0	0	7.1	0	0	3.0

10 mg of acid- and alkali-treated fraction IV-4 and 0.5 mg of pig pancreatic kallikrein were incubated in 2 ml of Britton-Robinson's universal buffer. The buffer was prepared from acid mixture (0.04 M phosphoric acid, 0.04 M acetic acid, and 0.04 M boric acid) adjusted to each desired pH value (3.0–10.0) using 0.2 M NaOH in the presence of angiotensinase inhibitors and 4×10^{-7} M SQ 14225, with or without 5.4×10^{-5} M pepstatin at 37 °C for 18 h. The vasoactive substances were extracted from the reaction mixture with hot ethanol and the extract evaporated to dryness under reduced pressure at 45 °C. Bioassays and radioimmunoassays for angiotensin I (AI-RI) and angiotensin II (AII-RI) in the products revealed that kallikrein generated pressor substance from acid- and alkali-treated fraction IV-4 at weakly acidic pH (4.0–6.0). Bioassay values of the pressor substance corresponded well with angiotensin II radioimmunoassay values. Pepstatin showed no inhibitory effect.

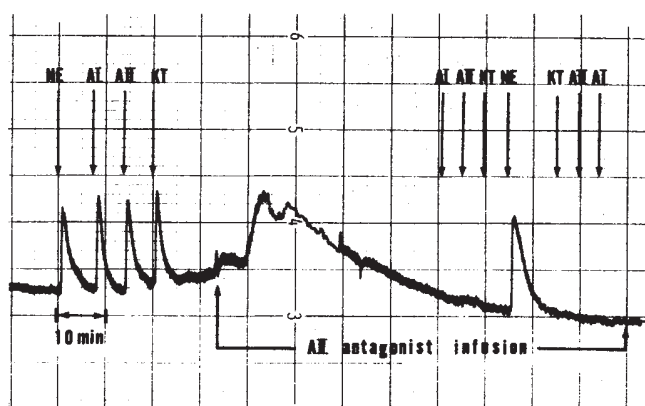


Fig. 1 Mode of pressor activity of kallikrein-generated pressor substance and effects of angiotensin II antagonist. Kallikrein-generated pressor substance was assayed by blood pressure in a female Wistar rat anaesthetized with intraperitoneal sodium amobarbital (100 mg per kg body weight) and ganglion-blocked with subcutaneous pentolinium tartrate (4 mg per kg body weight)¹⁴. Blood pressure was determined using a catheter inserted into the carotid artery by Hewlett-Packard pressure amplifier (Model 8805 C) and Yokogawa flat-recorder (Type 3051) connected to a Statham p23Db transducer. The samples were injected into a femoral vein, with [1-Sar,8-Ile]-angiotensin II injected into another femoral vein. By continuous infusion of [1-Sar,8-Ile]-angiotensin II at a rate of 50 ng per 0.05 ml per min, pressor activities of [5-Ile]-angiotensin I, [5-Ile]-angiotensin II, and kallikrein-generated pressor substance were inhibited, but that of noradrenaline was not. NE, 20 ng noradrenaline. AI, 6 ng angiotensin I. AII, 2 ng angiotensin II. KT, kallikrein-generated pressor substance.

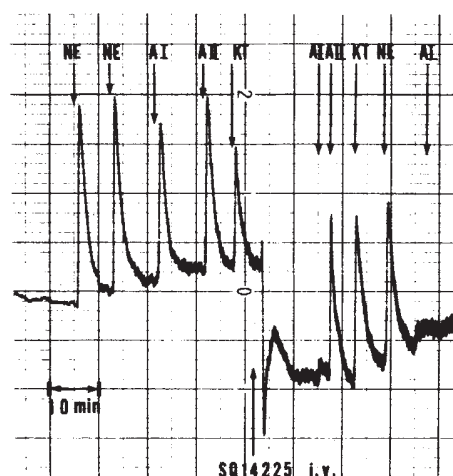


Fig. 2 Mode of pressor activity of kallikrein-generated pressor substance and effects of converting-enzyme inhibitor, SQ 14225. SQ 14225 was given by a bolus intravenous injection at a dose of 0.8 mg per kg body weight. It almost completely inhibited pressor action of [5-Ile]-angiotensin I, but not that of [5-Ile]-angiotensin II, kallikrein-generated pressor substance and noradrenaline. Symbols are the same as those in Fig. 1.

substances. Two opposing activities—depressor and pressor—are built into a one-enzyme system and the direction of the reaction may be determined by only a slight pH shift between 5.0 and 8.0. It is suggested that the term 'kinin(s)' be restricted to vasodepressor(s) and the term 'tensin(s)' to vasopressor(s). Based on this nomenclature, 'kinin-tensin enzyme system' may be coined for this type of one-enzyme system with bidirectional circulatory activities, such as the trypsin-(bradykinin and/or angiotensin II) system⁶, and the kallikrein-(kinin and/or angiotensin II) system.

We are attempting to purify and identify the pressor substance released by kallikrein *in vitro*, and examining the possibility that a comparable reaction occurs for the regulation of tissue perfusion *in vivo*. The fact that there must have been cleavage of the Phe-His bond by trypsin and kallikrein suggests that a more fundamental enzymological study with regard to their substrate specificity would be of interest.

This study was supported in part by grants-in-aid from the Japanese Ministry of Education (B-148195 and B-548200), the Fukuoka University Research Institute, and the Rinsho-yakuri-Kenkyushinko Foundation. We thank Miss M. Yanagiuchi and Y. Taniguchi for technical assistance.

Received 9 May; accepted 16 October 1980.

- Habermann, E. & Blennemann, G. *Naunyn-Schmiedeberg's Arch. exp. Path. Pharmacol.* **249**, 357–373 (1964).
- Elliot, D. F., Horton, E. W. & Lewis, G. P. *Biochem. J.* **78**, 60–65 (1961).
- Sealey, J. E., Atlas, S. A., Laragh, J. H., Oza, N. B. & Ryan, J. W. *Nature* **275**, 144–145 (1978).
- Yokosawa, N., Takahashi, N., Inagami, T. & Page, D. L. *Biochim. biophys. Acta* **569**, 211–219 (1976).
- Arakawa, K., Ikeda, M., Fukuyama, J. & Sakai, T. *J. clin. Endocr. Metab.* **42**, 599–602 (1976).
- Arakawa, K., Yuki, M. & Ikeda, M. *Biochem. J.* **187**, 647–653 (1980).
- Arakawa, K., Minohara, A., Yamada, J., Uemura, N. & Nakamura, M. *Clinica chim. Acta* **22**, 309–315 (1968).
- Habermann, E. *Handb. exp. Pharmacol.* **25**, 250–288 (1970).
- Arakawa, K., Minohara, A., Uemura, N., Sakai, T. & Ikeda, M. *Endocr. jap.* **22**, 472–432 (1975).
- Barret, A. J. *Biochem. J.* **104**, 601–608 (1967).
- Cushman, D. W. & Cheung, H. S. *Biochem. Pharmacol.* **20**, 1637–1648 (1971).
- Hummel, B. C. W. *Can. J. Biochem. Physiol.* **37**, 1393–1399 (1959).
- Aoyagi, T. *et al. J. Antibiotics* **25**, 689–694 (1972).
- Arakawa, K., Nakatani, M., Minohara, A. & Nakamura, M. *Biochem. J.* **104**, 900–906 (1967).

Are opioid peptides co-transmitters in noradrenergic vesicles of sympathetic nerves?

Steven P. Wilson*, Richard L. Klein†, Kwen-Jen Chang‡, Milton S. Gasparis†, O. Humberto Viveros* & Wen-Hsun Yang†

*Department of Medicinal Biochemistry, Wellcome Research Laboratories, Triangle Research Park, North Carolina 27709

† Department of Pharmacology and Toxicology, University of Mississippi Medical Center, Jackson, Mississippi 39216

‡ Department of Molecular Biology, Wellcome Research Laboratories, Research Triangle Park, North Carolina 27709

Biogenic amines and peptide hormones co-exist in paraneurons¹ or amine precursor uptake and decarboxylation (APUD) cells². Opioid peptide immunohistochemistry occurs in the gland cells and nerve endings of the adrenal medulla³, and both dopamine β -hydroxylase (EC 1.14.17.1) and enkephalin-like immunohistochemistry occurs in some small intensely fluorescent and sympathetic ganglion cells⁴. Direct biochemical assays demonstrate opiate-like peptides (OLPs) including Met- and Leu-enkephalin in the adrenal medulla, and the highest specific activity is attained in the purified adrenomedullary chromaffin granule fraction⁵. Chromaffin cells also synthesize enkephalins *de novo*⁶. Adrenal medullae from larger species, including ox and man, contain much more OLP material than those from smaller species such as rodents^{5,7}; fortunately so, because only from ox is it practical to purify both chromaffin granules and splenic nerve vesicles at particle purity and yield suitable for chemical composition studies. In ox and man the large dense-cored vesicles (LDVs) comprise 30–50% of the vesicle population in terminals of blood vessels, spleen and vas deferens, equivalent to 80–90% of the total terminal vesicle capacity. For these reasons, we chose bovine splenic nerve in which to investigate the localization of OLPs in the peripheral sympathetic system. Our results, reported here and based on direct biochemical analyses, indicate that OLPs, putative transmitters and noradrenaline probably co-exist in sympathetic C-fibres, which constitute 98% of the bovine splenic nerve. Further, the OLPs and noradrenaline can probably co-exist in the large dense-cored noradrenergic vesicles of this nerve. We believe the obtained 1:60 molar ratio of opioid peptides:noradrenaline to be the highest yet demonstrated for enkephalins in any neurotransmitter organelle or tissue.

The distribution of OLPs in the sucrose-D₂O density gradient used to purify noradrenergic LDVs is shown in Fig. 1. Of the original peptide activity layered onto the density gradient (SN₁₅), 14.6% is soluble and remains above the buffer zone. About 90% of the recoverable, sedimentable OLP material (Table 1) occurs in the heavy vesicle fractions (FIII, fractions 11 and 12) where LDVs equilibrate⁸. Negligible OLPs distribute in the less dense regions of the gradient, including the zone (fractions 6–8) where small dense-cored vesicles typically equilibrate.

The pattern of OLP distribution in the density gradient is almost identical to that of noradrenaline in sedimentable particles⁹, except that about one-third of the noradrenaline readily leaks from the vesicles during tissue homogenization. Dopamine β -hydroxylase also occurs at the highest specific activity in the heavy vesicle peak (Table 1), but about one-third is scattered throughout the gradient at lower buoyant densities⁹. This is explained by the occurrence of 'immature' LDVs from the

proximal end of the nerve, with full enzyme content but low noradrenaline content, and by depletion of matrix material resulting from vesicle rupture during isolation or from previous exocytosis.

Specific activities and recoveries of OLPs at various steps in the purification of LDVs are presented in Table 1 and compared with the analogous data for noradrenaline and dopamine β -hydroxylase. The molar ratio of noradrenaline:OLPs remains ~100:1 and the units of dopamine β -hydroxylase:nmol OLPs at ~1:1 throughout the purification process. The higher purification of OLPs over dopamine β -hydroxylase or noradrenaline (50-fold and 35-fold, respectively) is in keeping with its more restricted localization to LDVs and selective retention.

The stepwise recoveries of OLPs, dopamine β -hydroxylase and noradrenaline vary for particular reasons. Due to their molecular weight of ~500, OLPs should be released from the LDVs only by physical lysis or by exocytosis from nerve terminals, and therefore, relatively little should occur in the soluble phase of the nerve homogenate. However, the peptides are also readily degraded by ubiquitous peptidases¹⁰ and this could be partly responsible for the 25–35% loss reflected in recoverable peptide (Table 1 legend).

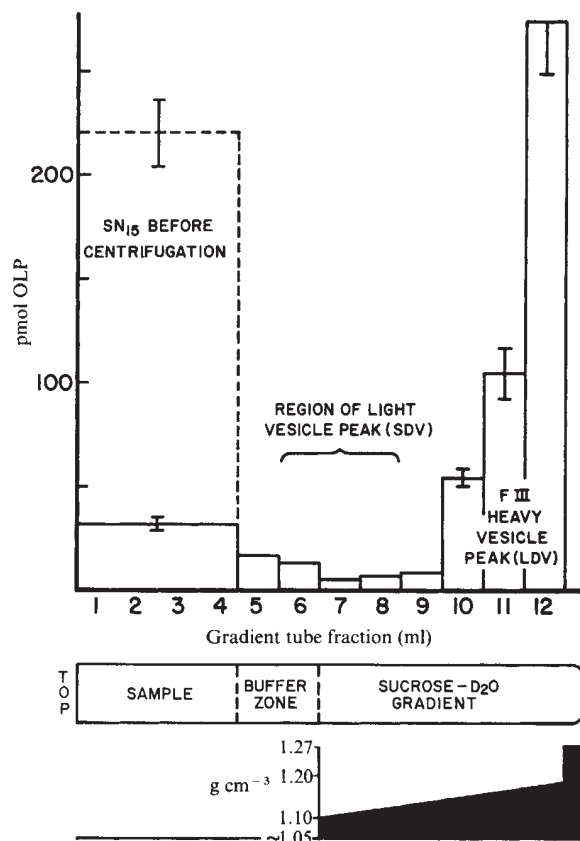


Fig. 1 Opiate-like peptide distribution following density gradient centrifugation. Purified LDVs were prepared from bovine splenic nerve^{8,12,13} obtained 10–15 min post mortem using a modified sucrose-D₂O density gradient sampling procedure⁹. OLPs were preserved in extracts containing 1.0 M acetic acid and measured by displacement of [¹²⁵I]-[D-Ala², D-Leu⁵]enkephalin (0.25 mM; specific activity 2 Ci μ mol⁻¹) from rat brain membrane receptors¹⁹. SN₁₅ is the supernatant obtained from the desheathed splenic nerve homogenate after centrifugation at 10,500g for 15 min. FIII is the purified heavy vesicle fraction containing LDVs and includes fractions (ml) 11 and 12 of the sucrose-D₂O density gradient based on their similar noradrenaline content per mg protein.

Table 1 Opiate-like peptide (OLP), noradrenaline (NA) and dopamine β -hydroxylase (D β H) content and recovery in purified vesicles from bovine splenic nerve

Fraction	NA (nmol per mg*† protein)	% Recovery per step	D β H*‡ (units per mg protein)§	% Recovery per step	OLPs (nmol per mg protein)	% Recovery per step	Protein (mg per gradient)¶
Whole nerve homogenate	1.0		0.011		0.10 $\pm$ 0.0008 (<i>n</i> = 12)		82–110
SN ₁₅ (sedimentable) (total)¶¶	7.1	67.8	0.024	57.6	0.020 $\pm$ 0.006 (<i>n</i> = 5)	51–63	27–31 62–73
FIII (fractions 11 and 12 of density gradient)	32.0	66.6	0.388	52.0	0.524 $\pm$ 0.043 (<i>n</i> = 11)	89–93*	0.7–1.09
Purification**	32x		35x		52x		
Overall recovery in gradient from SN ₁₅		~95		~100		69 (64–75)	

* Ref. 9.

† Ref. 13.

‡ M.S.G., W.-H.Y. and R.L.K., unpublished.

§ Unit of D β H activity equals 1 μ mol octopamine synthesized per min at 37 °C in a medium containing 20 mM tryamine as substrate.

¶ Protein distribution is based on the modified density gradient with 2.0 ml buffer zone.

¶¶ Percentage soluble OLPs in SN₁₅ (total) averaged 14.6% $\pm$ 1.0 (*n* = 12).* Recovery is corrected for 25.3–36.4% loss of OLPs which was due to degradation or other factors and is based on the actual overall recovery in the gradient from 4.0 ml SN₁₅.** Protein content per gradient and purification are uncorrected for nonspecifically adsorbed albumin content. OLP data are $\pm$ s.e.m.**Table 2** Distribution of opiate-like peptides in vesicles purified from sequential segments of bovine splenic nerve

Content	Proximal segment	Middle segment	Distal segment	% Increase
Per g nerve segment weight				
mg LDV protein	0.13 $\pm$ 0.02	0.16 $\pm$ 0.03	0.23 $\pm$ 0.01 (<i>P</i> < 0.01)	68%
nmol OLP	0.65 $\pm$ 0.06	0.80 $\pm$ 0.07	1.08 $\pm$ 0.12 (<i>P</i> < 0.02)	65%
units D β H*	0.047 $\pm$ 0.01	0.059 $\pm$ 0.01	0.083 $\pm$ 0.01 (<i>P</i> < 0.005)	77%
nmol ATP†	0.76 $\pm$ 0.02	—	1.32 $\pm$ 0.07 (<i>P</i> < 0.005)	74%
nmol NA‡	3.08 $\pm$ 0.20	5.13 $\pm$ 0.28	8.40 $\pm$ 0.17 (<i>P</i> < 0.001)	273%
Per mg LDV protein				
nmol OLP	0.74 $\pm$ 0.10	0.62 $\pm$ 0.16	0.64 $\pm$ 0.10	No change
units D β H*	0.62 $\pm$ 0.07	0.66 $\pm$ 0.08	0.64 $\pm$ 0.08	No change
nmol ATP†	10.0 $\pm$ 0.3	—	10.1 $\pm$ 0.5	No change
nmol NA‡	44.0 $\pm$ 2.6	57.0 $\pm$ 3.2	70.0 $\pm$ 4.0 (<i>P</i> < 0.005)	59%

Date are $\pm$ s.e.m., *n* = 4 paired experiments with duplicate OLP determinations in which vesicles were prepared identically and simultaneously from equal weights of desheathed splenic nerve segments: proximal and middle segments (extraspinal) were ~5–6 cm long and the distal segment (intrasplenic) was 15–30 cm long. See Table 1 legend for a definition of a unit of D β H activity.

* Ref. 9.

† Ref. 17.

‡ Refs 14, 18.

In the purified LDV fraction of the density gradient, FIII, the OLP content averages 0.5–0.6 nmol per mg protein. An additional twofold purification is to be expected for the LDVs by differential centrifugation of fraction FIII^{9,12,13}. This should result in OLP levels of ~1.0 nmol per mg protein, uncorrected for nonspecifically adsorbed bovine serum albumin, which amounts to at least one-third of the total protein in the vesicle fraction.

Analysis of purified LDV pellet extracts by HPLC reveals three major and several minor peaks with opioid activity (Fig. 2). Two of the major peaks, each representing 30% of the recovered opioid, have retention times identical to those of Met- and Leu-enkephalin standards. The ³H-Leu-enkephalin internal standard also eluted with a retention time of 19 min. The major opioid component with a retention time of 8 min probably corresponds to oxidized Met-enkephalin⁶. Recovery of opioid activity on the Amberlite column is 48%, compared with 80% for ³H-Leu-enkephalin internal standard, which is similar to recoveries in acetic acid extracts of adrenomedullary chromaffin

cells⁶. This probably represents the loss of opioid peptides of a larger molecular size than enkephalins¹¹. Recovery of OLPs eluted from the Amberlite column and subsequently subjected to HPLC is 70%, similar to the 60% recovery of ³H-Leu-enkephalin standard. Using the recovery of ³H-Leu-enkephalin to correct for losses of Met- and Leu-enkephalin during the chromatographic separations, these two peptides comprise 40–50% of the total opioid activity in LDV material.

Table 2 shows data from experiments designed to measure the levels of OLPs in the nerve vesicles as they are transported distally from the nerve cell bodies in the coeliac ganglion to the varicosities in the spleen. The amount of OLPs per gramme fresh nerve weight increases 60–70% in the distal direction and is equal to the increase in protein of the purified vesicle fraction, the latter being a measure of relative vesicle numbers per segment. Similar increases occur in dopamine β -hydroxylase activity and the ATP content, based on nerve segment fresh weight. Thus, OLPs, like dopamine β -hydroxylase and ATP, do not change significantly when expressed on the basis of total

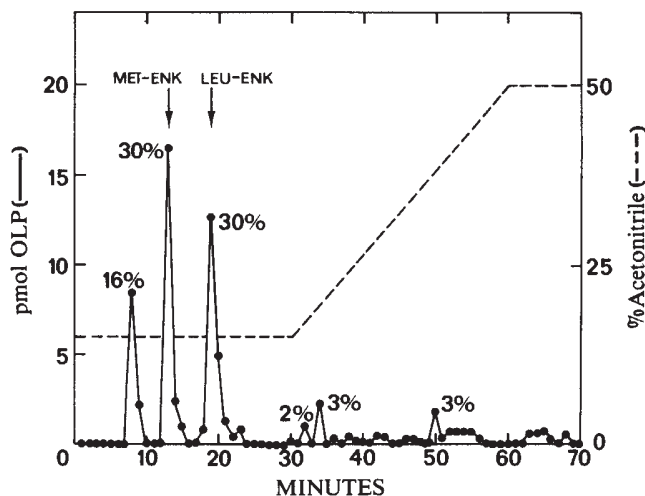


Fig. 2 High performance liquid chromatographic separation of opiate like peptides from bovine splenic nerve large dense-cored vesicles. An acid extract (2.6 ml) of LDV pellets was applied to a column of Amberlite XAD-2 (2 ml bed volume). The eluate was reappplied to the resin and the resin washed four times with 5 ml of 0.2 M acetic acid. Enkephalins were eluted from the column with 5 ml of 0.2 M acetic acid in methanol. The eluate was evaporated to dryness under N_2 and resuspended in water for application to the HPLC column. HPLC was carried out on a Partisil PXS-10, 10/25 ODS column. After sample application the column was eluted (1 ml min^{-1}) with 15% acetonitrile/10 mM ammonium acetate, pH 4.2, for 30 min followed by a gradient of 15–50% acetonitrile for 30 min. OLPs, opiate-like peptides; ENK, enkephalin.

protein in the purified LDV fraction from sequential nerve segments. In contrast, noradrenaline content per gram nerve fresh weight increases 273% in the sequential nerve segments. Thus, noradrenaline is synthesized and accumulated by the vesicles during axoplasmic transport, the content increasing ~60% per mg LDV protein¹⁴. This increase occurs exclusively in the fast release pool of 'newly synthesized' transmitter¹⁵.

This research was supported in part by grants NSF BNS-79-12951 and NIMH 1R03-MH35481 to R.L.K. and the AHA-Miss. Affil. and BRSG S07-RR05386 from the NIH to W.H.Y. Parts of this research have been reported elsewhere¹⁶.

Received 24 July; accepted 15 October 1980.

1. Fujita, T. & Kobayashi, S. *Trends neurol. Sci.* 27–30 (1979).
2. Pearse, A. G. E. *J. Histochem. Cytochem.* 17, 303–313 (1969).
3. Schultzberg, M. *et al. Neuroscience* 3, 1169–1186 (1978).
4. Schultzberg, M. *et al. Neuroscience* 4, 249–270 (1979).
5. Viveros, O. H., Diliberto, E. J., Hazum, E. & Chang, K.-J. *Molec. Pharmacol.* 16, 1101–1107 (1979).
6. Wilson, S. P., Chang, K.-J. & Viveros, O. H. *Proc. natn. Acad. Sci. U.S.A.* 77, 4364–4368 (1980).
7. Viveros, O. H., Diliberto, E. J., Hazum, E. & Chang, K.-J. *Adv. biochem. Psychopharmacol.* 22, 191–204 (1980).
8. Thureson-Klein, Å., Klein, R. L. & Lagercrantz, H. *J. Neurocytol.* 2, 13–27 (1973).
9. Klein, R. L. *et al. J. Neurobiol.* 10, 291–307 (1979).
10. Snyder, S. H. *Science* 209, 976–982 (1980).
11. Lewis, R. V. *et al. Science* 208, 1459–1461 (1980).
12. Lagercrantz, H., Klein, R. L. & Stjärne, L. *Life Sci.* 9, 639–650 (1970).
13. Yen, S. S., Klein, R. L. & Chen-Yen, S. H. *J. Neurocytol.* 2, 1–12 (1973).
14. Klein, R. L. in *Frontiers in Catecholamine Research* (eds Usdin, E. & Snyder, S.) 423–425 (Pergamon, Oxford, 1973).
15. Klein, R. L. & Harden, T. K. *Life Sci.* 16, 315–322 (1975).
16. Klein, R., Gasparis, M., Yang, W., Viveros, O. & Wilson, S. *Trans. Am. Soc. neurochem.* 11, 230 (1980).
17. Yen, S. S., Klein, R. L., Chen-Yen, S. H. & Thureson-Klein, Å. *J. Neurobiol.* 7, 11–22 (1976).
18. Chen-Yen, S. H. thesis, Univ. Mississippi (1975).
19. Miller, R. J., Chang, K.-J., Cooper, B. & Cuatrecasas, P. *J. biol. Chem.* 253, 531–538 (1978).

Differential regulation of the α_2 -adrenergic receptor by Na^+ and guanine nucleotides

Thomas Michel, Brian B. Hoffman & Robert J. Lefkowitz

Howard Hughes Medical Institute Laboratory, Departments of Medicine (Cardiology) and Biochemistry, Duke University Medical Center, Durham, North Carolina 27710

Many hormones interact with receptors which stimulate the enzyme adenylate cyclase. Less well characterized are those receptors which mediate an inhibition of adenylate cyclase activity¹. However, guanine nucleotides are clearly important in the regulation of both stimulatory and inhibitory receptors². Monovalent cations, notably Na^+ , regulate many inhibitory receptor systems but apparently not stimulatory receptors¹. We investigate here the effects of Na^+ and guanine nucleotides on the adenylate cyclase-coupled inhibitory α_2 -adrenergic receptor of the rabbit platelet³. Computer modelling of adrenaline competition curves with 3H -dihydroergocryptine (3H -DHE) indicates that adrenaline induces two distinct affinity states of the α_2 receptor—one of higher (α_{2H}) and the other of lower (α_{2L}) affinity. Guanylyl-5'-yl-imidodiphosphate (Gpp(NH)p) seems to reduce adrenaline affinity by converting the high-affinity state into the low-affinity form of the receptor. In contrast, Na^+ reduces adrenaline affinity at both the high- and low-affinity states of the α_2 receptor while preserving receptor heterogeneity. Thus, guanine nucleotides and Na^+ differ in the manner by which each reduces agonist affinity for the α_2 -adrenergic receptor.

The rabbit platelet's α receptors were found to be exclusively of the α_2 subtype by the computer modelling of competition curves of the antagonists prazosin and yohimbine with 3H -DHE⁴ which were best described by a model of one class of binding sites with yohimbine much more potent than prazosin (data not shown). This is similar to the situation in human platelets⁵.

Agonist interactions with the rabbit platelet's α_2 -adrenergic receptor are shown in Fig. 1 and Table 1. Without added Na^+ or guanine nucleotides, the adrenaline competition curve with 3H -DHE yielded a shallow binding isotherm with a slope factor ('pseudo' Hill coefficient) of 0.67 suggesting the presence of heterogeneous binding states. Computer modelling indicated the presence of two classes of binding states with high (α_{2H})- and low (α_{2L})-affinity states for adrenaline of K_d 6.3×10^{-9} M and 1.5×10^{-7} M, respectively. The addition of Gpp(NH)p (0.1 mM) caused the adrenaline competition curve to shift to lower affinity and to become steep with a slope factor of 1.0. These curves were best fitted by a model with one class of binding states of lower affinity ($K_d = 1.3 \times 10^{-7}$ M), with this affinity not significantly different from the lower affinity binding state ($K_{d\alpha_{2L}} = 1.5 \times 10^{-7}$ M) found in the absence of added Gpp(NH)p. These effects of Gpp(NH)p are very similar to those previously reported for the β -adrenergic receptor in frog erythrocytes⁶ and for α_2 -adrenergic receptors in human platelets⁵. In the absence of Gpp(NH)p, 57% of the α_2 receptors are in the α_{2H} state, the remainder are in the α_{2L} state, and the ratio $K_{d\alpha_{2L}}/K_{d\alpha_{2H}}$ is 24 (see Table 1). When 0.1 mM Gpp(NH)p is added, all of the α_2 receptors are present in the lower-affinity state (α_{2L}); there is no change in the total number of receptors present.

We have previously shown that 100 mM NaCl causes a maximal decrease in agonist affinity at these receptors⁷. In the presence of 100 mM NaCl, competition curves of (–)adrenaline and 3H -DHE are shifted to lower overall affinity and increased steepness relative to control curves (Fig. 1, Table 1). The observed slope factor, 0.8, is still, however, significantly less

Table 1 Parameters derived from computer modelling of competition curves of (–)adrenaline with ³H-DHE in rabbit platelet membranes in the presence and absence of NaCl and Gpp(NH)p

Additions	n	EC ₅₀ (M)	K _{dα_{2H}} (M)	K _{dα_{2L}} (M)	K _{dα_{2L}} /K _{dα_{2H}}		% R _H	% R _L	Slope factor
None*	3	7.2 ± 2.0 × 10 ⁻⁸	6.3 ± 1.0 × 10 ⁻⁹	1.5 ± 0.3 × 10 ⁻⁷	24	—	57 ± 8	43 ± 7	0.67 ± 0.07
0.1 mM Gpp(NH)p†	3	2.6 ± 0.8 × 10 ⁻⁷	—	1.3 ± 0.3 × 10 ⁻⁷ ‡	—	—	0	100	1.0 ± 0.10¶
100 mM NaCl*	5	2.5 ± 0.3 × 10 ⁻⁶	4.2 ± 0.5 × 10 ⁻⁷	3.2 ± 0.5 × 10 ⁻⁶	8	—	51 ± 20	49 ± 18	0.78 ± 0.09
0.1 mM Gpp(NH)p plus 100 mM NaCl†	4	4.2 ± 0.5 × 10 ⁻⁶	—	2.1 ± 0.5 × 10 ⁻⁶ §	—	—	0	100	1.1 ± 0.15¶

Binding assays were performed as described in the text. The parameters are derived from the computer modelling of the competition curves shown graphically in Fig. 1. The competition curves are the means of *n* experiments, each done in duplicate. EC₅₀ refers to the competitor concentration yielding half-maximal inhibition of ³H-DHE binding. K_{dα_{2H}} and K_{dα_{2L}} refer to the dissociation constants of the high- and low-affinity states, respectively. The fraction of α receptors in these states is shown as %R_H and %R_L, respectively. See text for details.

* Two-state fit significantly better than one-state fit (*P* < 0.01).

† Two-state fit not significantly better than one-state fit (*P* > 0.05).

‡ Value not significantly different from K_{dα_{2L}} in absence of NaCl and Gpp(NH)p (*P* > 0.05).

§ Value not significantly different from K_{dα_{2L}} in presence of NaCl without added Gpp(NH)p (*P* > 0.05).

|| Slope factor significantly different from 1.0 (*P* < 0.05).

¶ Slope factor not significantly different from 1.0 (*P* > 0.05).

than 1.0 (*P* < 0.05). The competition curve is significantly better fitted by a model with two classes of binding states for adrenaline. The proportion of the α₂-receptor population in the high (R_H)- and low (R_L)-affinity states is not significantly changed from the control curve, with %R_H = 51 and %R_L = 49. However, adrenaline affinity is reduced at both the high- and low-affinity states. There is a greater relative reduction in affinity at the high-affinity state, as shown by the ratio of K_{dα_{2L}}/K_{dα_{2H}} being 8 in the presence of 100 mM Na⁺ compared with a ratio of 24 in the absence of added Na⁺. These results indicate that both Na⁺ and Gpp(NH)p reduce agonist affinity, but that Gpp(NH)p abolishes the heterogeneity of agonist binding whereas receptor heterogeneity is maintained in the presence of Na⁺.

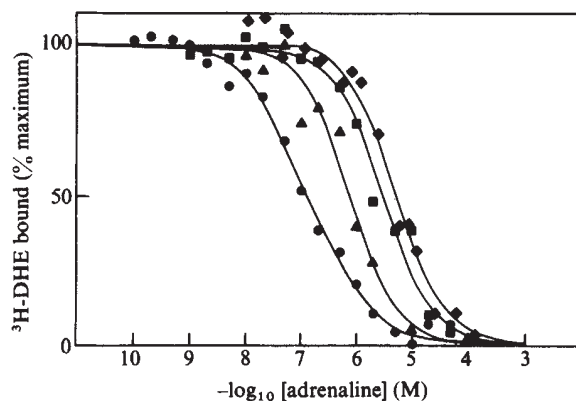
Addition of 0.1 mM Gpp(NH)p to adrenaline competition curves in the presence of 100 mM Na⁺ causes a further decrease in agonist affinity and an increase in the slope factor to 1.1 (Table 1; Fig. 1). The competition curves in the presence of Gpp(NH)p and Na⁺ are best described by a model with a homogeneous class of binding states. The affinity of this state for adrenaline (K_d = 2.1 × 10⁻⁶ M) is not significantly different from the K_{dα_{2L}} of 3.2 × 10⁻⁶ M determined in the presence of Na⁺ alone without added Gpp(NH)p. Thus, the effect of Gpp(NH)p in the presence

of Na⁺ is apparently to convert receptors in the α_{2H} state into the α_{2L} state, entirely analogous to the effect of this nucleotide on adrenaline competition curves in the absence of added Na⁺.

In rabbit platelet membranes, prostaglandin E₁ (PGE₁) causes a 10–12-fold increase in adenylate cyclase activity^{3,7}. In the absence of added NaCl, (–)adrenaline (0.1 mM) causes a small but significant decrease (11 ± 6%, data from five experiments) in PGE₁-stimulated (0.01 mM PGE₁) adenylate cyclase activity, from 169 ± 20 pmol cyclic AMP per min per mg protein to 150 ± 17 pmol cyclic AMP per min per mg protein. In the presence of 100 mM NaCl, there is a threefold greater maximal inhibition (31 ± 4%) of PGE₁-stimulated adenylate cyclase activity by adrenaline, from 207 ± 21 down to 143 ± 18 pmol cyclic AMP per min per mg protein (five experiments, significantly greater inhibition than in the absence of Na⁺, *P* < 0.005). The potentiation of adrenaline-mediated inhibition of PGE₁-stimulated adenylate cyclase by Na⁺ is dose dependent, with a half-maximal effect at a Na⁺ concentration of 30 mM (data not shown). In the conditions used in this assay, addition of GTP over a wide concentration range (10⁻⁷–10⁻⁴ M) had no apparent effect on the inhibition of adenylate cyclase by adrenaline in either the presence or absence of Na⁺.

Fig. 1 Competition curves of ³H-DHE with (–)adrenaline in the presence

and absence of Na⁺ and Gpp(NH)p. Isolation of rabbit platelet plasma membranes was as previously described¹⁰ with the exception that the lysate pellet was washed once in resuspension buffer containing 50 mM Tris-HCl and 10 mM MgCl₂ · 6H₂O, pH 7.5. The final resuspension was in this buffer to a protein concentration of 1–3 mg ml⁻¹. Radioligand binding assays were in a total volume of 0.2 ml which consisted of 0.025 ml of ³H-DHE (33 Ci mmol⁻¹, NEN) and 0.075 ml of drug in resuspension buffer with NaCl added to yield indicated final concentrations. Incubations (25 °C for 30 min) were started by addition of 0.1 ml of membranes. The assay was terminated by rapid vacuum filtration with a 20-ml wash with resuspension buffer at 25 °C. The concentration of ³H-DHE in these assays was 4 nM; specific radioligand binding to α receptors was 60–70% of total binding. The affinity of ³H-DHE binding was 3 × 10⁻⁹ M in all conditions used in these assays. Computer modelling of competition curves was done as described previously^{5,19–23}. For each treatment, competition curves were constructed with each point determined in duplicate. ³H-DHE binding, as a fraction of binding in the absence of competitor, was computed at each concentration of competitor. Then the results of the individual experiments were meaned and subjected to computer analysis. This analysis provided parameter estimates and standard errors which represent the confidence limits of the computer estimates. The computer program for the data analysis, which is based on the law of mass action, involves nonlinear least squares curve fitting²⁰ using a generalized model for complex ligand–receptor systems¹⁹. The computer technique provides an objective evaluation of whether the data are best fitted by a one- or two-receptor-state model. A fit of the data to two states of binding was accepted only if the fit was significantly improved (*P* < 0.05) over that for a one-state fit. The modelling provides values for the affinity constants of the competing ligand for the different states of the receptor and the proportion of the states present when a two-state fit is involved. Slope factors ('pseudo' Hill coefficients) and half-maximal inhibitory concentrations (EC₅₀s) in the various experimental conditions were determined using a four-parameter logistic equation, as previously described²². Adenylate cyclase assays were performed as described previously⁷ with slight modifications. The incubation mixture, in a final volume of 50 μl, consisted of 0.1 mM ATP, 1–2 × 10⁶ c.p.m. of [α-³²P]ATP and an ATP-regenerating system (in which none of the reagents is Na⁺ salts), 50 mM Tris-HCl pH 7.5, 10 mM MgCl₂, indicated concentrations of NaCl and 1–2 mg ml⁻¹ membrane protein. PGE₁ at 10⁻⁵ M caused an 8–12-fold stimulation of adenylate cyclase, with half-maximal stimulation by 10⁻⁷ M in either the presence or absence of NaCl. Protein was determined by the method of Lowry²⁴ using bovine serum albumin as standard. ●, No additions; ▲, +0.1 mM GppNHp; ■, +100 mM NaCl; ◆, +0.1 mM GppNHp + 100 mM NaCl.



The individual and additive reduction in agonist affinity promoted by Na^+ and guanine nucleotides is characteristic of diverse receptor systems which inhibit adenylate cyclase^{1,2,7-12}. Several studies have shown that the presence of one or both of these agents is required for the effective coupling of hormone binding to enzyme inhibition^{1,2,13-16}. The present experiments have demonstrated that α -adrenergic receptor-mediated inhibition of adenylate cyclase is potentiated by Na^+ . Note that the concentration of Na^+ yielding half-maximal reduction in agonist affinity (40 mM Na^+)⁷ is similar to the Na^+ concentration yielding half-maximal potentiation of adrenaline-mediated adenylate cyclase inhibition reported here. Our study did not demonstrate a GTP requirement for α -receptor-mediated adenylate cyclase inhibition. Presumably this reflects the presence of sufficient GTP in the adenylate cyclase assay used in this study to permit adenylate cyclase regulation in the absence of added guanine nucleotides.

These investigations have demonstrated heterogeneity of agonist binding to the α_{2L} and α_{2H} states of the α_2 -adrenergic receptor, with the α_{2H} state seemingly converted into the low-affinity form by guanine nucleotides. These observations resemble the situation described for the β -adrenergic receptor⁶, which promotes the GTP-dependent activation of adenylate cyclase. Applying both biochemical¹⁷ and computer-modelling techniques⁶⁻¹⁸ to the β -adrenergic receptor system, it has been proposed that heterogeneity of agonist binding and its regulation by guanine nucleotides are the result of the interaction of the receptor with a GTP-binding nucleotide regulatory protein in the membrane. It seems plausible that the α_2 -adrenergic receptor interacts with an analogous membrane-bound nucleotide regulatory protein¹³, and that this interaction is important in the guanine nucleotide-mediated modulation of agonist affinity states. We have shown that Na^+ does not abolish heterogeneity of agonist binding, and thus presumably the interaction of α receptor and nucleotide regulatory protein is maintained even in the presence of Na^+ as the addition of guanine nucleotides further reduces agonist affinity. In contrast to guanine nucleotides, Na^+ reduces affinity of both the high- and low-affinity states, and this effect of Na^+ may be independent of the nucleotide regulatory protein. Perhaps the action of Na^+ is mediated via the receptor itself or through an interaction with another membrane component distinct from the nucleotide regulatory protein.

In conclusion, these studies have indicated distinctly different roles for guanine nucleotides and Na^+ in the regulation of the adenylate cyclase-coupled inhibitory α_2 -adrenergic receptor. Although further studies are needed to establish the molecular basis of this differential regulation, the present data suggest that Na^+ effects may be mediated through independent mechanisms. Understanding the basis of the effects of Na^+ on either the receptor or other membrane components is relevant to the mechanism of hormone and drug inhibition of adenylate cyclase.

This work was supported by NIH grants HL16037 and HL20339. T.M. is a student in the Medical Scientist Training Program (NIH grant GM07171). B.B.H. is a Fellow of the MRC of Canada. R.J.L. is an Investigator of the Howard Hughes Medical Institute.

Received 28 May; accepted 11 August 1980.

- Jakobs, K. H. *Molec. cell. Endocr.* **16**, 147-156 (1979).
- Rodbell, M. *Nature* **284**, 17-22 (1980).
- Harwood, J. P., Moskowitz, J. & Krishna, G. *Biochem. biophys. Acta* **261**, 444-456 (1972).
- Hoffman, B. B., De Lean, A., Wood, C. L., Schocken, D. D. & Lefkowitz, R. J. *J. Life Sci.* **24**, 1739-1746 (1979).
- Hoffman, B. B., Mullikin-Kilpatrick, D. & Lefkowitz, R. J. *J. biol. Chem.* **255**, 4645-4652 (1980).
- Kent, R. S., De Lean, A. & Lefkowitz, R. J. *Molec. Pharmac.* **17**, 14-23 (1980).
- Tsai, B. S. & Lefkowitz, R. J. *Molec. Pharmac.* **14**, 540-548 (1978).
- Blume, A. J. *Proc. natn. Acad. Sci. U.S.A.* **75**, 1713-1717 (1978).
- Rosenberger, L. B., Yamamura, H. I. & Roeske, W. R. *J. biol. Chem.* **255**, 820-823 (1980).
- Tsai, B. S. & Lefkowitz, R. J. *Molec. Pharmac.* **16**, 61-68 (1979).
- Glossman, H. & Presek, P. *Naunyn-Schmiedeberg's Archs Pharmac.* **306**, 67-73 (1979).
- Greenberg, D. A., U'Prichard, D. C., Sheehan, P. & Snyder, S. H. *Brain Res.* **140**, 378-384 (1978).
- Steer, M. L. & Wood, A. J. *J. biol. Chem.* **254**, 10791-10797 (1979).
- Blume, A. J., Lichtenstein, D. & Boone, G. *Proc. natn. Acad. Sci. U.S.A.* **76**, 5626-5630 (1979).

- Lichtenstein, D., Boone, G. & Blume, A. J. *Cyclic Nucleotide Res.* **5**, 367-375 (1979).
- Aktories, K., Schultz, G. & Jakobs, K. H. *FEBS Lett.* **107**, 100-104 (1979).
- Limbird, L. E., Gill, D. M. & Lefkowitz, R. J. *Proc. natn. Acad. Sci. U.S.A.* **77**, 775-779 (1980).
- Stadel, J. M., De Lean, A. & Lefkowitz, R. J. *J. biol. Chem.* **255**, 1436-1441 (1980).
- Feldman, H. A. *Analyt. Biochem.* **48**, 317-338 (1972).
- Fletcher, J. E. & Schrager, R. I. *A User's Guide to Least Squares Curve Fitting Tech. Rep.* No. 1 (U.S. Dep. of Health, Education and Welfare, 1973).
- Rodbard, D. & Frazier, G. R. *Meth. Enzym.* **37B**, 3-22 (1975).
- De Lean, A., Munson, P. J. & Rodbard, D. *Am. J. Physiol.* **235**, E97-E102 (1978).
- Hancock, A. A., De Lean, A. & Lefkowitz, R. J. *Molec. Pharmac.* **16**, 1-9 (1979).
- Lowry, O. H., Rosebrough, N. J., Fain, A. L. & Randall, R. J. *J. biol. Chem.* **193**, 265-275 (1951).

No junctional communication between epithelial cells in hydra

S. W. de Laat*, L. G. J. Tertoolen* & C. J. P. Grimmelikhuijzen†‡

* Hubrecht Laboratory, International Embryological Institute, Uppsalalaan 8, 3584 CT Utrecht, The Netherlands

† Max-Planck-Institut für medizinische Forschung, Abteilung Biophysik, Jahnstrasse 29, 6900 Heidelberg, FRG

Diffusion gradients of morphogens have been inferred as a basis for the control of morphogenesis in hydra^{1,2}, and morphogenetic substances have been found which, on the basis of their molecular weight (MW), should be able to pass gap junctions³⁻⁷. There have been several reports of the presence of gap junctions between epithelial cells of hydra⁸⁻¹³. However, until now, there has been no report published on whether these junctions enable the epithelial cells to exchange molecules of small molecular weight, as has been described in other organisms¹⁴⁻²⁰. Therefore we decided to investigate the communicative properties of the junctional membranes by electrophysiological methods and by intracellular-dye iontophoresis. We report here that no electrotonic coupling is detectable between epithelial cells of *Hydra attenuata* in: (1) intact animals, (2) head-regenerating animals, (3) cell re-aggregates, and (4) hydra that have become nerve-free. Furthermore we show that epithelial cells are unable to exchange low-molecular weight fluorescent dyes.

Electrotonic coupling between epithelial cells was measured by conventional intracellular glass microelectrodes²¹. Current pulses were passed and voltages recorded through a single microelectrode. In the intact animal measurements were performed in the head, mid-gastric and foot regions with the two electrodes positioned at short distances (<100 μm) along the longitudinal or latitudinal axis of the body column. The chances of impalement of epithelial cells were high as this cell type covers most of the surface of the animal¹¹. In addition to normal animals, we tested for electrotonic coupling in animals in which we could expect special morphogenetic regulation. In head-regenerating animals electrotonic coupling was monitored at regular intervals in the new head region, starting 30 min after head removal and continuing until new tentacles had appeared (2 days). In another experiment we dissociated hydra into its separate component cells and allowed these to reaggregate, sort, and form new animals²². We made measurements all over the aggregates from 2 h after the onset of aggregation until new heads appeared (3 days). Phenotypic variants of hydra that consisted almost totally of epithelial cells were also measured. These so-called nerve-free hydra have similar gap junctions to normal animals¹³ and are useful in microelectrode studies as they do not contract spontaneously and offer better chances for impaling epithelial cells only.

By careful impalement with the microelectrodes it was possible to distinguish in intact animals between the ectodermal and the endodermal cells. Figure 1 shows an example of a recording in which the measurement was initially taken from an ectodermal cell, and further penetration resulted in the impalement of an underlying endodermal cell. Ectodermal cells have

‡ To whom correspondence should be addressed.

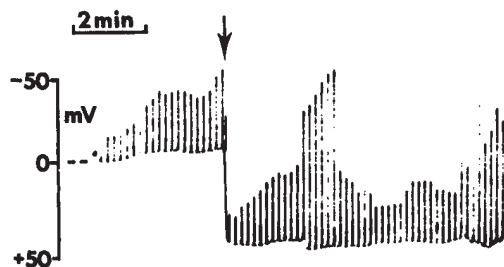


Fig. 1 Example of an intracellular recording of the membrane potential and input resistance from epithelial cells of intact normal hydra. Initially an ectodermal cell was impaled, but on further penetration (arrow) the microelectrode moved into an underlying endodermal cell. The recording demonstrates the occurrence of spontaneous fluctuations in membrane potential and input resistance, and the difference in membrane potential between ectodermal and endodermal cells. The input resistance was determined as the ratio between the deflection of the membrane potential and the amplitude of the current pulse. Intracellular voltages were measured and current pulses (5×10^{-10} A, 500 ms) were passed through a single 3 M KCl-filled glass microelectrode by the use of a bridge circuit to compensate for the electrode impedance. Hydra was kept in a medium composed of 1 mM CaCl_2 , 0.1 mM MgCl_2 , 0.1 mM KCl, 0.5 mM Na_2HPO_4 , adjusted to pH 7.6 with phosphoric acid. Microelectrode tip potentials (<10 mV) and resistances (20–40 M Ω) were determined and compensated in this medium. The electrode properties depend on the ionic strength of the medium. In a control experiment we measured a decrease in tip potential $<+5$ mV and electrode resistance <15 M Ω upon changing to a 305 mosm salt solution (about equivalent to the cytoplasm). For this reason membrane potentials and input resistances are expected to be 5 mV more negative and 15 M Ω greater, respectively, than measured. The mobility of the intact and the head-regenerating animals was reduced during the measurements by adding 2% of the polymer Tylose (Hoechst) to the medium. This allowed continuous recordings of up to 15 min. With nerve-free animals and cell aggregates no Tylose was used.

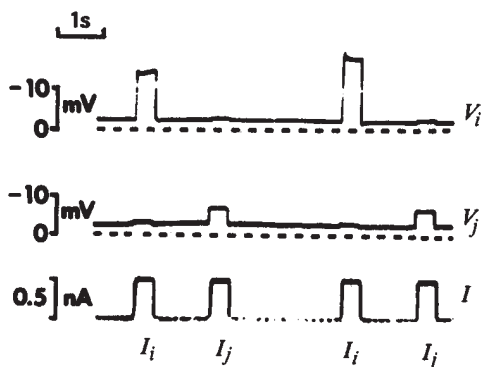


Fig. 2 Lack of electrotonic coupling between visibly adjacent ectoderm cells in nerve-free hydra. The figure gives an example of the maximum deflection in membrane potential that could be observed on passing current into an adjacent cell—in nearly all measurements no visible deflection could be observed. In all experiments described in the text the coupling ratio V_i/V_j was <0.04 , where V_i and V_j are the deflection in the membrane potential of cell i and cell j , respectively, when a current pulse is passed from cell j to the grounded medium. Nerve-free hydra were obtained as described¹³ and consisted of more than 95% epithelial cells. Conditions as in Fig. 1.

membrane potentials ranging between +5 and -15 mV, whereas endodermal cells have membrane potentials of +25 to +50 mV. This was confirmed by histological examination of fixed animals after marking the impaled cells with Lucifer Yellow CH. The difference in membrane potential between both types of epithelial cells agrees with earlier observations by Kass-Simon and Diesl on isolated epithelial cells²³. In both types of cell we found spontaneous fluctuations in membrane potential and input resistance (Fig. 1). As no significant differences in membrane potentials and input resistances were found between intact, nerve-free and head-regenerating animals, these data were pooled (Table 1). In early re-aggregates we could not distinguish between ectodermal and endodermal cells. Here the membrane potentials ranged between -15 and +25 mV, with mean value $+9.1 \pm 5.3$ mV (s.e.m.), $N = 7$. The input resistance ranged between 35 and 100 M Ω , with a mean value of 68.6 ± 7.6 M Ω , $N = 7$ (see also Fig. 1).

Over 100 measurements failed to show electrotonic coupling among ectodermal cells, among endodermal cells, or between ectodermal and endodermal cells in all the experimental conditions, including those in which special morphogenetic regulation

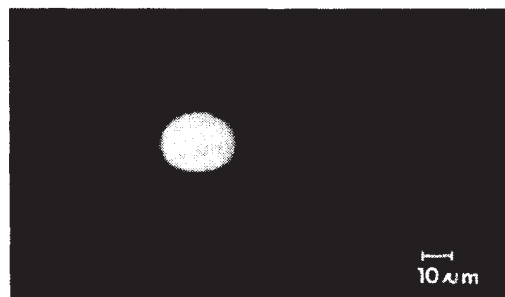


Fig. 3 Example of a micrograph taken 30 min after iontophoresis of Lucifer Yellow CH into an epithelial cell in nerve-free hydra. No transfer of dye to other cells was observed in any of the experimental conditions applied. Similar results were obtained with fluorescein. The tip of a regular microelectrode was filled with a 3% aqueous solution of Lucifer Yellow CH, whereas the remainder of the electrode system was filled with 3 M LiCl. The dye was introduced iontophoretically into the cell by applying repetitive hyperpolarizing current pulses of 2×10^{-9} A and 1.5 s duration at 0.5 s intervals for 1 min. Fluorescence was observed and photomicrographs were taken as described²⁷.

might be expected. Taking into account the resolution limit of our registration system, the coupling ratio V_i/V_j (see Fig. 2 legend) was always smaller than 0.04. An example of a measurement in nerve-free hydra is shown in Fig. 2.

It might be argued that possible leakage of calcium from the medium into the cells during the impalement might have caused a closure of the gap junctions²⁰. To check this, we repeated the measurements in calcium-free medium. We also determined the possible effect of the trans-epithelial potential which normally exists between the gastric cavity and the external medium^{24,25}. By insertion of a capillary into the mouth this potential was abolished. In all these circumstances we found normal membrane potentials, input resistances and absence of electrotonic coupling.

In addition to our electrophysiological experiments we studied the exchange of Lucifer Yellow CH (MW, 457) and fluorescein (MW, 330) between epithelial cells. Lucifer Yellow CH^{26,27} particularly, was extremely effective for tracing dye coupling due to its excellent spectral properties, its rapid spread through an injected cell, and good retention by the cell during histological processing. In over 75 experiments fluorescent dye was introduced iontophoretically from a glass microelectrode into epithelial cells at different positions along the body column of intact and nerve-free animals. The localization of the dye was

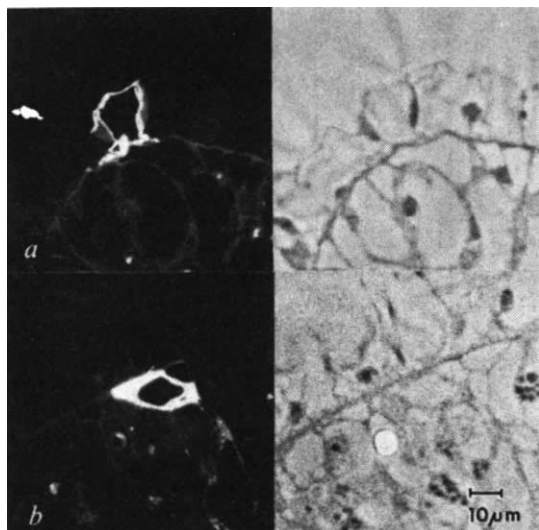


Fig. 4 Micrographs of 1- μ m histological sections of normal hydra fixed after iontophoresis of Lucifer Yellow into an ectodermal cell (a) and endodermal cell (b); left: fluorescence, right: phase contrast (unstained). Conditions as in Fig. 3. The animals were fixed in 4% formaldehyde for 24 h at 4°C, dehydrated in dimethoxypropanal (DMP) for 1 h, embedded in Spurr low-viscosity embedding medium by placing in Spurr/DMP (1:1) for 5 h, Spurr/DMP (3:1) for 12 h and Spurr only for 8 h. Polymerization was at 60°C for 12 h.

monitored by fluorescence microscopy up to 1 h after application. Throughout this time the fluorescence remained definitively located within a single cell, without detectable transfer to neighbouring cells (Fig. 3). Some of the animals were fixed after introduction of Lucifer Yellow CH into a number of epithelial cells and examined after histological processing (Fig. 4). From observation of serial 1- μ m sections we made the following conclusions: (1) there was no transfer of the dye between epithelial cells, (2) the absence of electrotonic and dye coupling was not due to positioning of the micro-electrode in the large vacuoles—iontophoretically applied Lucifer Yellow CH was definitively located within the cytoplasm, (3) the impalement did not cause apparent cell damage, (4) ectodermal and endodermal cells have different membrane potentials (see above).

Our observations show that the epithelial cells in hydra are not able to exchange ions and fluorescent dyes of low molecular weight. We conclude that: (1) gap junctions between epithelial cells of hydra, observed by electron microscopy, do not function as communicative intercellular junctions—they might be of a primitive type, involved in the formation of connective junctions, (2) morphogenetic substances probably diffuse via the extracellular space, (3) nerve-free hydra which are still able to contract on electrical or mechanical stimulation²⁸ cannot conduct the contraction pulse along electrotonically coupled epithelial cells—they may use paracrine communication.

We thank Dr W. W. Stewart for his gift of Lucifer Yellow CH, Dr H. C. Schaller for her interest, and the Deutsche Forschungsgemeinschaft (Scha 253/6) for financial support.

Table 1 Membrane potential and input resistance in hydra epithelial cells

	Membrane potential (mV)	Input resistance (M Ω)
Ectoderm	-2.5 ± 1.2 (N = 16)	38.8 ± 4.7 (N = 16)
Endoderm	$+36.0 \pm 1.3$ (N = 40)	45.5 ± 2.9 (N = 40)

Mean number of cells $\pm$ s.e.m. are given. The spontaneous fluctuations in each cell (see Fig. 1) were averaged by taking values of the membrane potential and the input resistance at 1-min intervals.

Received 23 July; accepted 16 October 1980.

- Gierer, A. & Meinhardt, H. *Kybernetik* **12**, 30–39 (1972).
- Wolpert, L., Clarke, M. R. B. & Hornbruch, A. *Nature new Biol.* **239**, 101–105 (1972).
- Schaller, H. C. *J. Embryol. exp. Morph.* **29**, 27–38 (1973).
- Berking, S. *Wilhelm Roux's Arch. dev. Biol.* **181**, 215–225 (1977).
- Grimmelikhuijzen, C. J. P. *Cell Differentiation* **8**, 267–273 (1979).
- Grimmelikhuijzen, C. J. P. & Schaller, H. C. *Trends biochem. Sci.* **4**, 265–267 (1979).
- Schmidt, T. & Schaller, H. C. *Wilhelm Roux's Arch. dev. Biol.* **188**, 133–139 (1980).
- Hand, A. R. & Gobel, S. J. *J. Cell Biol.* **52**, 397–408 (1972).
- Filshie, B. K. & Flower, N. E. *J. Cell Sci.* **23**, 151–172 (1977).
- Wood, R. L. *J. ultrastruct. Res.* **58**, 299–315 (1977).
- West, D. L. *Tissue Cell* **10**, 629–646 (1978).
- Wood, R. L. *Cell Tissue Res.* **199**, 319–338 (1979).
- Lepault, J., McDowall, A. W. & Grimmelikhuijzen, C. J. P. *Cell Tissue Res.* **209**, 217–224 (1980).
- Gilula, N. B., Reeves, O. R. & Steinbach, A. *Nature* **235**, 262–265 (1972).
- Bennett, M. V. L. *Fedn Proc.* **32**, 65–75 (1973).
- Azarnia, R., Larsen, W. J. & Loewenstein, W. R. *Proc. natn. Acad. Sci. U.S.A.* **71**, 880–884 (1974).
- Johnson, R., Hammer, M., Sheridan, J. D. & Revel, J. P. *Proc. natn. Acad. Sci. U.S.A.* **71**, 4536–4540 (1974).
- Pitts, J. D. & Simms, J. W. *Expl Cell Res.* **104**, 153–163 (1977).
- Loewenstein, W. R. in *Differentiation and Development* (eds Ahmed, F., Russell, T. R., Schultz, J. & Werner, J.) 399–409 (Academic, New York, 1978).
- Loewenstein, W. R. *Biochim. biophys. Acta* **560**, 1–65 (1979).
- de Laat, S. W., Barts, P. W. J. A. & Bakker, M. I. *J. Membrane Biol.* **27**, 109–129 (1976).
- Gierer, A. *et al. Nature new Biol.* **239**, 98–101 (1972).
- Kass-Simon, G. & Diesl, V. K. *Nature* **265**, 75–77 (1977).
- Josephson, R. K. & Macklin, M. *Science* **156**, 1629–1631 (1967).
- Josephson, R. K. & Macklin, M. *J. gen. Physiol.* **53**, 638–665 (1969).
- Stewart, W. W. *Cell* **14**, 741–759 (1978).
- de Laat, S. W., Tertoolen, L. G. J., Dorresteyn, A. W. C. & van den Biggelaar, J. A. M. *Nature* **287**, 546–548 (1980).
- Campbell, R. D., Josephson, R. K., Swab, W. E. & Rushforth, N. B. *Nature* **262**, 388–390 (1976).

Receptor binding and internalization of immobilized transcobalamin II by mouse leukaemia cells

Kiyoshi Takahashi*, Mehdi Tavassoli* & Donald W. Jacobsen†‡

Departments of *Clinical Research and †Biochemistry, Scripps Clinic and Research Foundation, La Jolla, California 92037

Membrane transport of vitamin B₁₂ (cyanocobalamin; Cbl) into mammalian cells is mediated by the serum protein transcobalamin II (TCII)^{1–3}. In mouse leukaemia L1210 cells, TCII–Cbl binds to membrane receptors in a rapid, temperature-independent step and is internalized by a slow, temperature-dependent process⁴. To delineate the location of receptors on these cells, we have constructed a visual probe by covalently coupling purified TCII–Cbl to submicrometre latex particles⁵ (minibeads). We report here that when L1210 cells are incubated with minibeads containing TCII–Cbl at 4°C and examined by scanning electron microscopy (SEM), the particles are found attached predominantly to microvilli. Incubation of the cells at 37°C results in the internalization of the minibeads. As visualized by transmission electron microscopy (TEM), this endocytotic process seems to occur in clathrin-coated pits and vesicles^{6,7} at the cell surface.

Recent work has demonstrated that cultured L1210 cells attach tightly to large (70 μ m) Sephacryl beads coated with Cbl–TCII⁸. Although the binding of numerous cells (~150 cells per bead) somewhat obscured the contact zone between the bead and the cell, it was apparent that cellular microvilli made intimate and extensive contact with the bead surface. To identify specific receptor areas for TCII–Cbl on L1210 cell surfaces and perhaps confirm the apparent role of microvilli in the transport process, we have repeated these experiments using small (0.35 μ m) latex particles derivatized with TCII–Cbl. The latter were prepared by a method modified from Weston and

‡ To whom correspondence should be addressed.

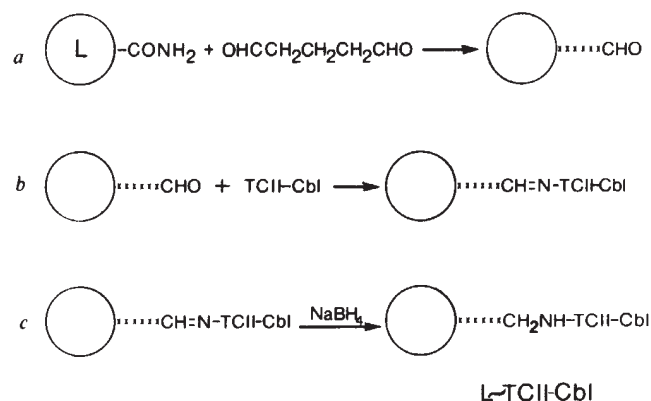


Fig. 1 Synthesis and structure of the receptor probe L~TCII-Cbl. Amide-modified polystyrene latex particles (0.35 μm ; Dow) were suspended to 5% (w/v) in 0.1 M NaHCO_3 (pH 8.0) and 8% glutaraldehyde was added to a final concentration of 5% (step a). After 24 h at 37 °C, excess glutaraldehyde was removed by dialysis against 0.1 M NaHCO_3 (pH 8.8) for 24 h at 4 °C. Rabbit serum TCII-Cbl (100 μg), purified by the method of D.W.J. *et al.*¹⁰, was added to 1 ml of a 0.1% suspension of the glutaraldehyde-modified minibeads in 0.1 M NaHCO_3 (pH 8.8) (step b). After 6 h at room temperature, the preparation was treated with 0.2 M NaBH_4 for 15 min at 4 °C (step c). The L~TCII-Cbl minibeads were pelleted by centrifugation (40,000g for 20 min) and washed three times with phosphate-buffered saline (pH 7.4).

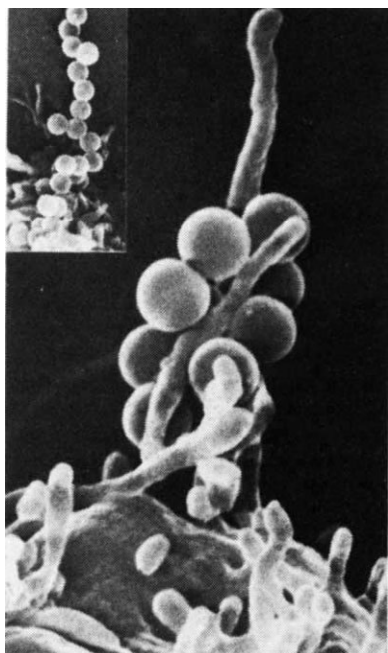


Fig. 2 Binding of L~TCII-Cbl to microvilli of L1210 cells. Cultured L1210 cells were suspended for 2 h (4 °C) with 0.04% L~TCII-Cbl in RPMI 1640 medium containing 5% fetal bovine serum, collected by centrifugation (150g for 7 min) and washed twice with serum-free medium. A drop of cell suspension was placed on a round coverslip pretreated with poly-L-lysine (50 μg ml^{-1}) and fixed in modified Karnovsky's solution containing 0.125 M cacodylate buffer (pH 6.8) for 7 min. The coverslip was then rinsed three times with 0.125 M cacodylate (pH 7.3) and postfixed in similarly buffered 2% OsO_4 for 15 min at room temperature. After dehydration and displacement in Freon 113, the cells were critical point-dried and sputter coated with platinum-palladium. $\times 18,000$, inset $\times 4,500$.

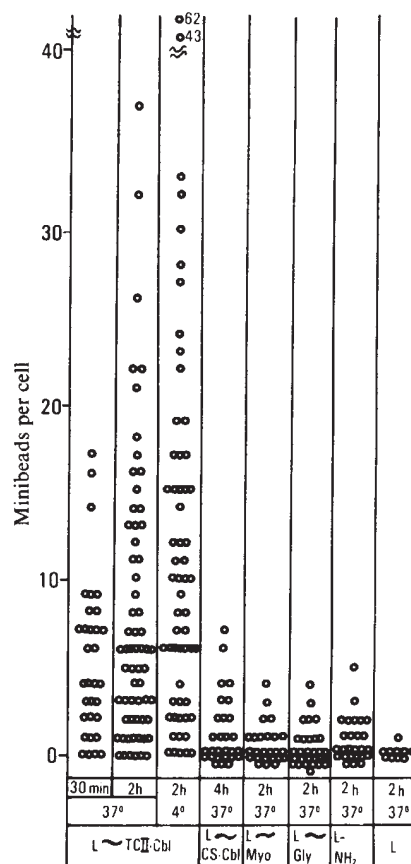


Fig. 3 Binding capacity of L1210 cells for L~TCII-Cbl and control minibeads. Incubation conditions were as described in Fig. 2 legend except that time and temperature varied as indicated. Each circle represents the number of minibeads bound per cell. Scoring was made on SEM prints of cells randomly selected at low magnification ($\times 500$; well below the resolution required to detect minibeads) and then photographed at higher magnification ($\times 6,000$). These data represent two separate preparations of L~TCII-Cbl. Control studies used plain (L) or amide-modified (L-NH₂) latex, or minibeads on which TCII-Cbl was replaced by glycine (L~Gly), myoglobin (L~Myo) or the non-TCII Cbl-binder from chicken serum (L~CS-Cbl).

Avrameas⁹ to form the receptor probe L~TCII-Cbl shown in Fig. 1. The same method was used to prepare control minibeads derivatized with glycine, myoglobin or the non-TCII Cbl-binder from chicken serum.

Cultured L1210 cells were incubated with L~TCII-Cbl at 4 or 37 °C for 0.5, 2 or 4 h. The cells were then collected, washed, fixed and processed for SEM and TEM. As shown in Fig. 2, SEM revealed extensive interaction between L~TCII-Cbl and microvilli. In some cases, these minibeads decorated the entire length of the microvilli, suggesting a localized and highly concentrated region of TCII-Cbl receptors. Minibead clusters were occasionally seen closer to the cell surface, but shorter microvilli could have been responsible for this effect. Quantitative data (Fig. 3) obtained from SEM prints indicated that minibeads containing TCII-Cbl bound equally well at 4 or 37 °C, averaging 12 and 10 per cell, respectively. The specificity of the interaction is also shown in Fig. 3. Thus, unmodified minibeads or minibeads derivatized with glycine, myoglobin or the non-TCII Cbl-binder from chicken serum did not bind to L1210 cells.

When cells were incubated with L~TCII-Cbl at 37 °C and then examined in sections using TEM, various stages of internalization were visualized (Fig. 4). Small clusters of minibeads appeared at the cell surface and were embedded in coated pits

and coated vesicles. The latter were morphologically similar to vesicles containing the structural protein clathrin, but were somewhat larger ($\sim 0.35 \mu\text{m}$) than those commonly encountered in adsorptive endocytosis ($0.05\text{--}0.25 \mu\text{m}$)⁶.

These studies provide visual evidence for the localization of TCII-Cbl receptors on the microvilli of L1210 cells and suggest that the complex is internalized at the surface of the cell. Although it is not clear how the receptor complex reaches the surface for endocytosis, at least two mechanisms can be invoked: either the microvilli retract and become contiguous with the surface, or the receptor complex laterally diffuses along the

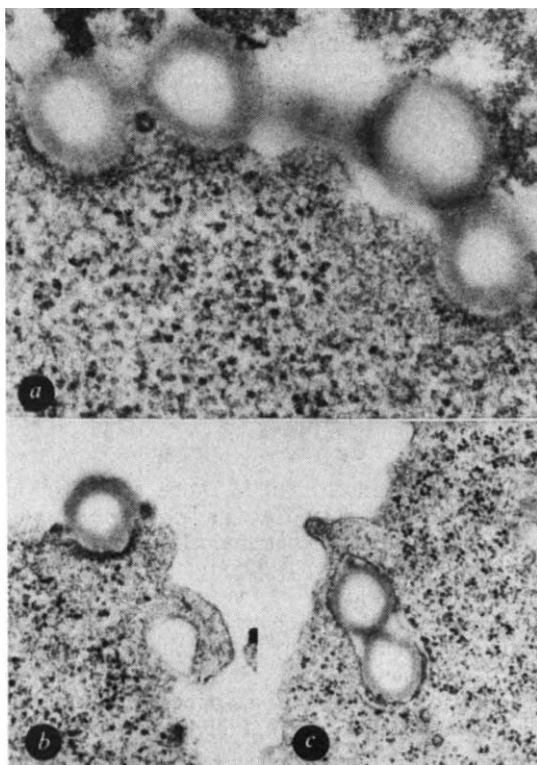


Fig. 4 Internalization of L-TCII-Cbl by L1210 cells. Incubation conditions were as described in Fig. 2 legend except that the temperature was 37°C . Cells were prepared, fixed and dehydrated as for SEM, and then embedded in Spurr resin and thin-sectioned. After staining with uranyl acetate and lead citrate, preparations were examined by TEM. *a* Shows several minibeads on the cell surface, two of which seem to be in the process of internalization. $\times 21,000$. *b* Shows a minibead within a vesicle which seems to be coated by clathrin. $\times 15,000$. *c* Shows two minibeads completely internalized within a common vesicle. $\times 10,000$.

microvilli and on to the surface for subsequent endocytosis. In either case, internalization of L-TCII-Cbl seems to occur via clathrin-coated pits and vesicles in a temperature-dependent fashion.

This work was supported by NIH grants AM-25510 (M.T.), AM-25406 (D.W.J.), CA-16600, DOE contract EV00899 (M.T.) and American Cancer Society grant CH-31.

Received 22 July; accepted 16 October 1980.

- Cooper, B. A. & Paranchych, W. *Nature* **191**, 393-395 (1961).
- Mahoney, M. J. & Rosenberg, L. E. in *Cobalamin* (ed. Babior, B. M.) 369-402 (Wiley-Interscience, New York, 1975).
- Youngdahl-Turner, P., Mellman, I. S., Allen, R. H. & Rosenberg, L. E. *Exp Cell Res* **118**, 127-134 (1979).
- DiGirolamo, P. M. & Huennekens, F. M. *Archs Biochem. Biophys.* **168**, 386-393 (1975).
- Molday, R. S., Dreyer, W. J., Rembaum, A. & Yen, S. P. S. *J. Cell Biol.* **64**, 75-88 (1975).
- Pearse, B. *Trends biochem. Sci.* **5**, 131-134 (1980).
- Goldstein, J. L., Anderson, R. G. W. & Brown, M. S. *Nature* **279**, 679-685 (1979).
- Jacobsen, D. W., Montejano, Y. D., Vitols, K. S. & Huennekens, F. M. *Blood* **55**, 160-163 (1980).
- Weston, P. D. & Avrameas, S. *Biochem. biophys. Res. Commun.* **45**, 1574-1580 (1971).
- Jacobsen, D. W. & Montejano, Y. D. in *Vitamin B12* (eds Zagalak, B. & Friedrich, W.) 947-954 (de Gruyter, Berlin, 1979).

Growth-modulating plasma tripeptide may function by facilitating copper uptake into cells

Loren Pickart*, Jonathan H. Freedman†, W. John Loker‡, Jack Peisach†, Christopher M. Perkins‡, Ronald E. Stenkamp§ & Boris Weinstein‡

* Virginia Mason Research Center, Seattle, Washington 98101

† Department of Molecular Pharmacology, Albert Einstein College of Medicine, Bronx, New York 10461

‡ Department of Chemistry and § Department of Biological Structure, University of Washington, Seattle, Washington 98195

The plasma tripeptide glycyl-L-histidyl-L-lysine (GHL), when added at nanomolar concentrations to a wide group of cultured systems, produces a disparate set of responses ranging from the stimulation of growth and differentiation to outright toxicity¹⁻³. Such diverse actions imply that this tripeptide mediates some basic biochemical function common to many types of cells and organisms. During the isolation of GHL we found the compound to co-isolate through a number of steps with approximately equimolar copper and about 1/5 molar iron¹. Maximal effects on hepatoma cells (HTC₄) were seen when the peptide was added with copper and iron to the growth medium¹. Structure-function studies revealed that several tripeptides with a histidyl-lysyl linkage were nearly as active as GHL¹. The association of GHL with copper and a homology similarity between the tripeptide and the copper transport sites on albumin and α -fetoprotein, where the cupric atom is bound to a histidyl residue adjacent to a basic residue, suggested that GHL may act as a copper transport factor⁴. We report here that the tripeptide readily forms complexes with copper(II) and enhances the uptake of the metal into cultured hepatoma cells.

GHL was synthesized by coupling N^α -t-butyloxycarbonyl- N^{im} -benzyloxycarbonyl-L-histidine to benzyl N^ϵ -benzyloxycarbonyl-L-lysine using a mixed anhydride procedure involving isobutyl chloroformate and N -methyl morpholine to yield benzyl N^α -t-butyloxycarbonyl- N^{im} -benzyloxycarbonyl-L-histidyl- N^ϵ -benzyloxycarbonyl-L-lysinate. The protected dipeptide was dissolved in methanol saturated with HCl to cleave the t-butyloxycarbonyl protecting group and to precipitate benzyl N^{im} -benzyloxycarbonyl-L-histidyl- N^ϵ -benzyloxycarbonyl-L-lysinate hydrochloride. Neutralization of this salt with triethyl amine gave the free dipeptide amine, which on joining to N^α -benzyloxycarbonyl-glycine by a mixed anhydride method, formed benzyl N^α -benzyloxycarbonyl-glycyl- N^{im} -benzyloxycarbonyl-L-histidyl- N^ϵ -benzyloxycarbonyl-L-lysinate. Hydrogenolysis of this fully protected compound produced the free peptide and chromatography on silica gel yielded crystalline GHL that had satisfactory microanalytical and amino acid analyses, as well as spectral data (IR and NMR).

When GHL was added to hepatoma (HTC₄) cells, which had been cultured in low serum (0.75%) fetal calf serum, it was observed that the tripeptide elevated copper uptake in a dose-dependent manner from 10 to 200 ng ml⁻¹. Higher concentrations of the compound were less effective and very high concentrations of the tripeptide ($10 \mu\text{g ml}^{-1}$) were inhibitory to copper uptake into the cells. When the compound was added at 200 ng ml⁻¹, copper uptake increased approximately 2.4-fold after 4 h incubation (Fig. 1). The tripeptide's effect was rapid and the rise in the intake of copper at 30 min was 50% of the 4-h value. Addition of the constituent amino acids of GHL at a molar ratio 300-fold greater than that found in the parent compound at 200 ng ml⁻¹ did not alter the rate of copper transfer (Fig. 1). By contrast, the tripeptide had no significant effect on the transport of ferrous (⁵⁹Fe) iron into the cells. These

results suggest that GHL transports copper, and perhaps other related metals, into cells in culture.

This possibility is further supported by evidence that the tripeptide readily forms chelates with cupric ions. A single crystal X-ray diffraction study of a GHL-Cu complex showed the existence of a tridentate bonding structure involving the N-terminal group of glycine, the nitrogen atom of the first amide bond, and the imine nitrogen atom of the imidazole ring of histidine. The ϵ -terminal amine group on the lysine side chain remains free and is not bound to the copper atom (Fig. 2). The electron paramagnetic resonance spectrum (EPR) of the peptide was determined on a frozen, aqueous solution (0.55 mM) in the presence of Cu(II) (0.50 mM) at 77 K and 9.073 GHz. The magnetic parameters obtained from the axial spectrum were $A_{||}$ 210 G, $g_{||}$ 2.335, and $g_{\perp}$ 2.057. These data indicate that the $d_{x^2-y^2}$ orbital is the highest d-orbital, a result consistent with strong interactions between a copper ion and four ligands in a near-equatorial arrangement. The super-hyperfine structure present in the EPR spectrum supported the involvement of the copper with at least one nitrogen donor atom in a near-planar coordination. Electron spin echo studies confirmed the coordination of imidazole to copper and there was no indication from the EPR spectrum for dipolar interactions between two adjacent copper atoms⁵. The combined X-ray and EPR results are consistent with the idea of a GHL-Cu complex in solution that is monomeric in nature.

These experiments at low-serum conditions of cell culture suggest an increase in copper transport serves as a mechanism for the action of GHL and may explain the range of responses to

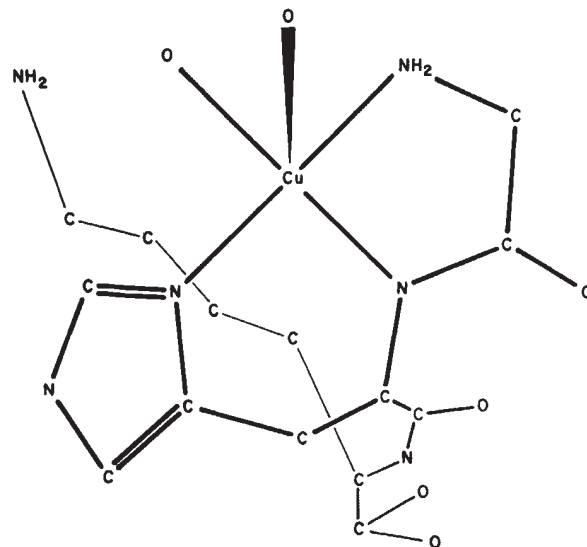


Fig. 2 Structure of GHL-Cu crystals. X-ray analysis indicates that the N-terminal group of glycine, the adjacent nitrogen in the first amide linkage of the peptide chain and the unprotonated nitrogen of the imidazole ring of histidine form bonds to copper. EPR measurements suggest that the GHL-Cu complex exists as a monomeric species in dilute solution. The detailed interpretation of the X-ray and EPR data will be published elsewhere.

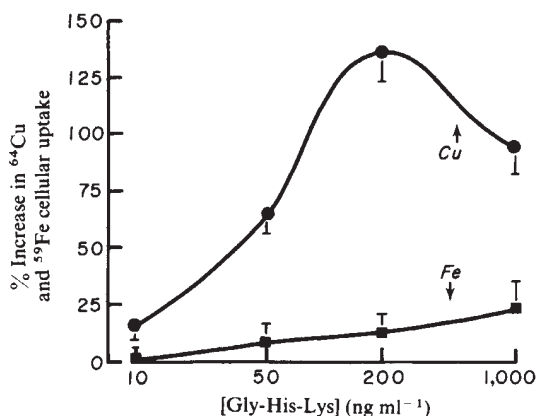


Fig. 1 Effect of glycy-L-histidyl-L-lysine (GHL) on the uptake of ⁶⁴Cu and ⁵⁹Fe into hepatoma cells in culture. Hepatoma (HTC4) cells were cultured for 3 days in 90% Eagle's basal medium plus 10% Swim's S-77 medium and 0.7% fetal calf serum in plastic T-flasks (25 cm², 10⁶ cells per flask) by methods previously detailed¹. GHL was then added to the cultures at the indicated dosages. ⁶⁴Cu (1 μ Ci, 4.9 Ci per g) was added and glycine, histidine and lysine were present in 300-fold molar excess over GHL at 200 ng ml⁻¹. The cells were incubated for 4 h at 37 °C, the media removed, and the cells washed four times with phosphate buffer saline, pH 7.4 (Gibco), at 0 °C, incubated in 0.01% trypsin for 15 min, and then cells were removed from the flask. The hepatoma cells were centrifuged at 2,000g for 30 min at 0 °C then homogenized in distilled water. The cellular debris was removed by centrifugation at 10,000g for 10 min and aliquots of the supernatant, representing copper from the cellular cytosol, counted by normal scintillation counting methods²¹. The cellular uptake of radiolabelled Cu at the various dosages of GHL ($\pm$ s.d.) for six experiments is given in Fig. 1. In the complete absence of serum, GHL was only one-quarter as active in stimulating Cu uptake. The control Cu uptakes at 0.5, 1 and 2 h were, respectively, 51%, 72% and 84% of the 4-h value. The uptake of radiolabelled Fe (1 μ Ci per flask, 6.2 Ci per g) was determined in the same way as for Cu.

the compound. Synthetic tripeptide, when added to culture media at concentrations of 10–200 ng ml⁻¹, has been found to: (1) enhance the growth of hepatoma cells³, thyroid follicular cells⁶ and T-strain mycoplasma⁷; (2) stimulate the growth and differentiation of *Ascaris* larvae⁸, lymphocytes⁹ and neurones^{10,11}; (3) assist the establishment in culture of 19 cancer cell lines from human tumours¹²; (4) raise the viability of eosinophils¹³, endometrial cells¹⁴, fibroblasts¹⁵, kidney cells¹⁶, liver cell and organ cultures^{3,17}, macrophages¹⁸, mast cells^{19,20} and placental organ cultures²¹; (5) inhibit the growth of glial¹⁰ and L929 cells²²; and (6) increase antibody cytotoxicity towards parasitic worms (*Schistosoma mansoni*)²³. Different patterns of cellular response to an increased copper(II) intake would rationalize the range of biological effects observed by various workers.

Furthermore, these results clearly delineate a structural model for a molecule that delivers transition metals to cellular receptors. The first two residues of the GHL molecule are involved in the bonding of the copper, whereas the side chain of lysine may be involved in the recognition of receptors that function in the uptake of the copper into cells. This possibility is further strengthened by the obvious analogies to the copper transport sites on plasma albumin or, in the fetus, on α -fetoprotein, in which the histidyl residue that binds the copper is immediately adjacent to either a lysyl or an arginyl residue.⁴ These considerations imply that the combination of a histidyl unit next to a basic residue constitutes a biologically active structure critical for the uptake of copper into cells.

We thank Drs Lyle H. Jensen and Norman J. Rose for their help and advice, and Madams Barbara Charles and Ruth Rehbein for the typing. The work was supported by USPHS grants AM 17702, CA 28858, GM 10828, HL 13399 and RR 00558.

Received 30 July; accepted 16 October 1980.

- Pickart, L. & Thaler, M. *Nature new Biol.* **243**, 85–87 (1973); *J. Chromat.* **175**, 65–73 (1979); *FEBS Lett.* **104**, 119–122 (1979); *J. cell Physiol.* **102**, 129–139 (1980).
- Schlesinger, D. H., Pickart, L. & Thaler, M. *Experientia* **33**, 324–325 (1977).
- Pickart, L., Thayer, L. & Thaler, M. *Biochem. biophys. Res. Commun.* **54**, 562–566 (1973).
- Aoyagi, T., Ikenaka, T. & Ichida, F. *Cancer Res.* **38**, 3483–3486 (1978).
- Peisach, J. in *Mechanisms of Oxidizing Enzymes* (eds Singer, T. P. & Ondarza, R. N.) 385–406 (Elsevier, Amsterdam, 1978).
- Ambesi-Impimbatto, F. S., Parks, L. A. M. & Coon, H. G. *Proc. natn. Acad. Sci. U.S.A.* **77**, 3455–3459 (1980).

7. Robertson, J. A. *J. clin. Microbiol.* **7**, 127–132 (1978).
8. Stromberg, B. E., Khoury, P. B. & Soulsby, E. J. L. *Int. J. Parasit.* **7**, 149–151 (1977).
9. Dessaint, J. P., Camus, D., Fischer, E. & Capron, A. *Eur. J. Immun.* **7**, 624–629 (1977).
10. Sensenbrenner, M., Jaros, G. G., Moonen, G. & Mandel, P. *Neurobiology* **5**, 207–213 (1975).
11. Lindner, G., Grosse, G. & Henklein, P. Z. *mikrosk.-anat. Forsch.* **93**, 820–828 (1979).
12. Simon, W. E. & Holzel, F. J. *Cancer Res. clin. Oncol.* **94**, 307–323 (1979).
13. Capron, M., Capron, A., Torpier, G., Bazin, H., Bout, D. & Joseph, M. *Eur. J. Immun.* **8**, 127–133 (1978).
14. Svanberg, L. & Astedt, B. *Experientia* **35**, 818–819 (1979).
15. Leung, M. K., Fessler, L. J., Greenberg, D. B. & Fessler, J. H. *J. biol. Chem.* **254**, 224–232 (1979).
16. Astedt, B., Barlow, G. & Holmberg, L. *Thromb. Res.* **11**, 149–153 (1977).
17. Eriksson, S., Alm, R. & Astedt, B. *Biochim. biophys. Acta* **542**, 496–505 (1978).
18. Joseph, M., Dessaint, J. P. & Capron, A. *Cell Immun.* **34**, 247–258 (1977).
19. Capron, M., Rousseaux, J., Mazingue, C., Bazin, H. & Capron, A. *J. Immun.* **121**, 2518–2526 (1978).
20. Mazingue, C., Dessaint, J. P. & Capron, A. *J. immun. Meth.* **21**, 65–77 (1978).
21. Holmberg, L., Lecander, I., Persson, B. & Astedt, B. *Biochim. biophys. Acta* **544**, 128–137 (1978).
22. Slotta, K. H., Golub, A. L. & Lopez, V. *Hoppe-Seyler's Z. physiol. Chem.* **356**, 367–376 (1975).
23. Torpier, G., Quaiisi, M. A. & Capron, A. *J. ultrastruct. Res.* **67**, 276–287 (1979).

Identification of the BAL-labile factor

E. C. Slater & S. de Vries

Laboratory of Biochemistry, BCP Jansen Institute, University of Amsterdam, Plantage Muidergracht 12, 1018 TV Amsterdam, The Netherlands

One of us has previously reported¹ that treatment of the Keilin and Hartree heart-muscle preparation² with 2,3-dimercaptopropanol (BAL), in the presence of air, leads to the complete inactivation of the succinate oxidase system with little if any effect on the activities of succinate dehydrogenase (until more than half the BAL was oxidized) or cytochrome *c* oxidase. The inactivation of the complete succinate oxidase system requires the oxidation of BAL by air in the presence of the enzyme. It is not caused by H₂O₂ or BAL disulphides produced during the oxidation of BAL. Spectroscopic studies identified the block as lying between cytochromes *b* and *c*. It was suggested that a BAL-labile factor is present which transfers electrons from cytochrome *b* to cytochrome *c* and which is destroyed by coupled oxidation with BAL. The factor is also required for NADH oxidation³. Subsequent work showed that it is not identical with cytochrome *c*₁ (ref. 4), myoglobin present in the preparation⁵ or the antimycin-binding site⁶. We report here that this factor is identical to the iron-sulphur protein in the central portion of the respiratory chain first identified by Rieske⁶.

The following clues led us to this identification: (1) Deul and Thorn's observation⁵ that, whereas after either BAL treatment or addition of antimycin electron flow between cytochromes *b* and *c* is blocked, the addition of antimycin to BAL-treated preparations causes a block in the reduction of cytochrome *b*. (2) Trumpower's observation⁷ that in preparations of succinate/cytochrome *c* oxidoreductase from which Racker's 'oxidation factor'⁸ has been extracted electron transfer between cytochromes *b* and *c*₁ is impaired, but that in the presence of antimycin the reduction of cytochrome *b* is blocked. (3) The recent identification by Trumpower⁹ of the active principle in Racker's preparations of oxidation factor as Rieske's iron-sulphur protein.

Direct evidence that BAL specifically destroys the Rieske iron-sulphur protein has now been obtained by electron paramagnetic resonance (EPR) spectrometry.

The EPR spectrum of sub-mitochondrial particles reduced with ascorbate in the presence of tetramethyl-*p*-phenylenediamine, cytochrome *c* and cyanide (Fig. 1) shows lines characteristic of the Rieske iron-sulphur protein: $g_z = 2.03$, $g_y = 1.89$ and $g_x = 1.80$. There is an additional line at about 1.94 derived from partial reduction of one of the iron-sulphur clusters associated with succinate dehydrogenase, and a free radical

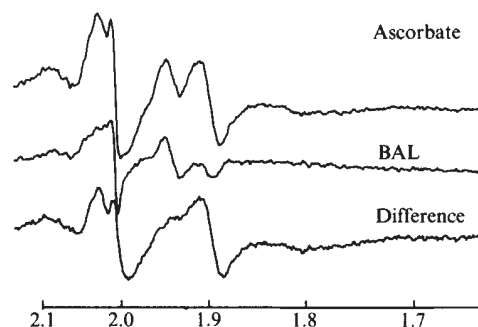


Fig. 1 EPR spectrum of sub-mitochondrial particles (70 mg ml⁻¹ in 0.25 M sucrose, 50 mM morpholinopropane sulphonic acid buffer, pH 7.0) reduced for 3 min at 20 °C with 5 mM ascorbate in the presence of 10 μM *N,N,N',N'*-tetramethyl-*p*-phenylenediamine, 20 μM cytochrome *c* and 4 mM KCN. The EPR spectra were measured at 36 K, 5 mW microwave power, 9.33 GHz frequency and 6.3 G modulation amplitude. The spectra plotted were 20–25 times scan averaged and corrected for differences in tube diameter and protein concentration. The top spectrum is that of untreated particles and the middle one after incubation with 10 mM BAL (2,3-dimercaptopropanol) and shaking in air for 30 min at 36 °C (ref. 11). The bottom spectrum is the difference: control minus BAL-treated.

at about $g = 2.0$ derived from the semiquinone of ubiquinone and from ascorbate. (The broad line at about 2.1 is probably due to contaminating copper.) After BAL treatment, the signals of the Rieske iron-sulphur protein completely disappeared. The residual signals are due to the free radical and the g_y and g_z lines of the iron-sulphur cluster of succinate dehydrogenase. The difference spectrum shows the three lines characteristic of the Rieske iron-sulphur protein.

Figures 2 and 3 show similar spectra obtained after reduction with succinate and NADH + Na₂S₂O₄, respectively. Although the g_z and g_y lines are now overshadowed by signals from other iron-sulphur clusters reduced at the lower potentials, the difference spectra (control minus BAL-treated) are in both cases dominated by the signals of the Rieske cluster. The complete elimination of the broad g_x line at 1.80, which is characteristic of the Rieske iron-sulphur protein, is particularly clear. Lines at $g = 1.92$ and 2.01 in the difference spectrum in Fig. 3 are

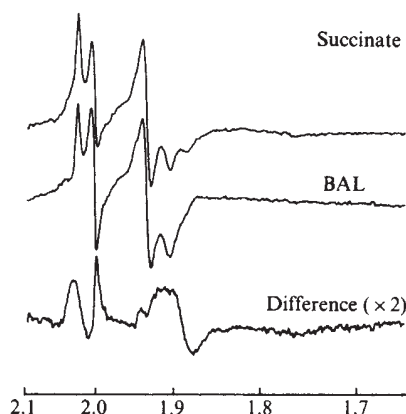


Fig. 2 EPR spectrum of sub-mitochondrial particles reduced for 10 min at 30 °C with 20 mM succinate in the presence of 20 μM cytochrome *c* and 4 mM KCN. Conditions were as in Fig. 1 legend, except that the gain used for the direct spectra was 1/1.6 times that in Fig. 1, and that in the difference spectrum was twice that in the direct spectrum.

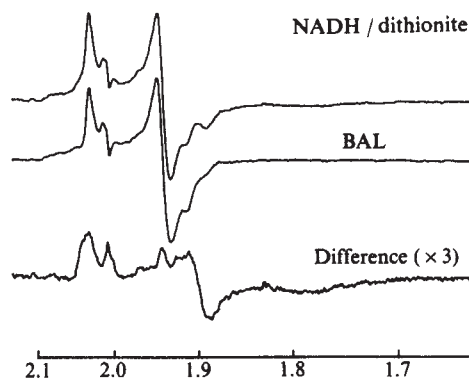


Fig. 3 EPR spectrum of sub-mitochondrial particles reduced for 10 min at 20 °C with 2 mM NADH (pH 7.2) and dithionite in the presence of 20 μ M cytochrome *c* and 4 mM KCN. Conditions were as in Fig. 1 legend, except that the gain used for the direct spectra was 1/3.2 times that in Fig. 1, and that in the difference spectrum was three times that in the direct spectrum.

probably due to destruction by BAL treatment of an iron-sulphur protein in the outer membrane of the mitochondrion, fragments of which are present in sub-mitochondrial particles¹⁰.

Figure 3 shows that, although the mitochondrial inner membrane contains about 10 different iron-sulphur clusters, only that belonging to the Rieske protein is destroyed by BAL

treatment of sub-mitochondrial particles, thus confirming the specificity of BAL treatment¹¹.

The nature of the chemical reaction between BAL, oxygen and the iron-sulphur protein is not established. It is possible that a disulphide bridge is formed between BAL and one of the cysteine molecules or the inorganic sulphur that is bound to the iron atom in the iron-sulphur cluster. The specificity of the BAL action may be due to the relative accessibility of this strongly hydrophilic compound to the Rieske iron-sulphur cluster.

We conclude that the BAL-labile factor is identical to the Rieske iron-sulphur protein; the oxidation factor has also been identified with this protein⁹. BAL treatment is therefore a useful technique for specifically eliminating the Rieske iron-sulphur cluster from the respiratory chain of sub-mitochondrial particles.

This work was supported in part by a grant from the Netherlands Organization for the Advancement of Pure Research (ZWO) under the auspices of the Netherlands Foundation for Chemical Research (SON).

Received 21 August; accepted 4 November 1980.

1. Slater, E. C. *Nature* **161**, 405–407 (1948).
2. Keilin, D. & Hartree, E. F. *Proc. R. Soc. B* **129**, 277–306 (1940).
3. Slater, E. C. *Nature* **165**, 674–676 (1950).
4. Slater, E. C. *Nature* **163**, 532 (1949).
5. Deul, D. H. & Thorn, M. B. *Biochim. biophys. Acta* **59**, 426–436 (1962).
6. Rieske, J. S., Hansen, R. E. & Zaugg, W. S. *J. biol. Chem.* **239**, 3017–3022 (1964).
7. Trumpower, B. L. *Biochem. biophys. Res. Commun.* **70**, 73–80 (1976).
8. Nishibayashi-Yamashita, H., Cunningham, C. & Racker, E. *J. biol. Chem.* **247**, 698–704 (1972).
9. Trumpower, B. L. & Edwards, C. A. *J. biol. Chem.* **254**, 8697–8706 (1979).
10. Heidrich, H. G., Albracht, S. P. J. & Bäckström, D. *FEBS Lett.* **95**, 314–318 (1978).
11. Slater, E. C. *Biochem. J.* **45**, 14–30 (1949).

1,25-Dihydroxycholecalciferol stimulation of a mitochondrial protein in chick intestinal cells

Adrian N. Hobden*, Marilyn Harding
& D. Eric M. Lawson

Dunn Nutritional Laboratory, University of Cambridge and Medical Research Council, Milton Road, Cambridge CB4 1XJ, UK

The steroid hormone 1,25-dihydroxycholecalciferol (1,25-(OH)₂D₃) stimulates the absorption of dietary calcium by the small intestine of animals although the exact mechanism by which this is achieved remains unclear. However, it has long been known that a soluble, calcium-binding protein (CaBP), is produced in large amounts in the cytoplasm of the intestinal cells of animals after *in vivo* administration of vitamin D₃ or 1,25-(OH)₂D₃ (refs 1, 2). We report here that 1,25-(OH)₂D₃ administered *in vivo* to rachitic chickens also stimulates production of another protein with molecular weight (MW) 39,000–42,000 which is insoluble in the absence of detergent, is found in the outer mitochondrial membrane and is produced in advance of maximum calcium transport.

Synthesis of CaBP is not optimal until after maximum stimulation of calcium transport by 1,25-(OH)₂D₃ (ref. 3)—there is significant stimulation of calcium transport before any CaBP can be detected in the intestinal cell. Thus CaBP cannot have a direct effect on calcium transport and other factors which are responsive to *in vivo* administration of 1,25-(OH)₂D₃ have been sought. Two such factors, proteins of MW 45,000 and 84,000, have been identified in the intestinal brush borders of chickens^{4,5}. However, only small quantities of these proteins

were produced after administration of 1,25-(OH)₂D₃ possibly indicating additional rather than *de novo* synthesis. The function of these proteins is unknown although any role for them in calcium uptake would possibly be at the level of absorption of calcium from the intestinal lumen into the cytoplasm. Once calcium has crossed the intestinal brush border membrane, it has to be transported across the cell to the basal membranes and then into the capillaries. We therefore looked for any changes in protein composition of intestinal mitochondria following *in vivo* administration of 1,25-(OH)₂D₃ to rachitic chickens.

At various times following administration of 125 ng of 1,25-(OH)₂D₃, the rachitic chickens were killed and slices of their everted jejunum incubated in medium containing [¹⁴C]leucine. Mitochondria were isolated from the slices and their proteins electrophoresed through SDS gels together with proteins (labelled with [4,5-³H]leucine) from the intestinal mitochondria of 1,25-(OH)₂D₃-deficient chickens. The gels were sliced and the radioactivity in each slice determined. Examination of the ratio of ¹⁴C:³H in each slice revealed that an extra protein (or additional amounts of a protein) was being produced in the birds subjected to 1,25-(OH)₂D₃ administration (Fig. 1). The level of synthesis of this protein was quantified from the double-label plot⁶ (Fig. 1) and a time course of synthesis determined (Fig. 2). This additional mitochondrial protein maximally accounted for over 7% of total mitochondrial protein synthesis (Fig. 2). Comparison with the time courses for other events which follow *in vivo* administration of 1,25-(OH)₂D₃ (refs 3, 4) reveals that maximum production of this mitochondrial protein (at 4 h) occurred before maximum calcium transport (at 8 h) and well before maximum synthesis of CaBP (Fig. 3). Although the MW and the time course of appearance (Fig. 3) were similar to those of a protein previously observed in intestinal brush borders^{4,5} the mitochondrial protein was produced in greater amounts (P. Wilson, personal communication), had a consistently smaller MW (39,000–42,000) and did not behave like actin (putatively identified as the brush border protein⁷) on gel electrophoresis in 8 M urea. We suggest therefore that the mitochondrial protein is a novel product of the

* To whom correspondence should be addressed.

intestinal cells which is produced following *in vivo* administration of $1,25-(\text{OH})_2\text{D}_3$.

Partial fractionation of the intestinal mitochondria into inner membrane, outer membrane and soluble protein fractions was achieved by sonication and sucrose density-gradient centrifugation⁷. In this way we showed that the novel mitochondrial protein had a distribution between the fractions similar to the enzyme rotenone-insensitive NADH-cytochrome *c* reductase (Fig. 4), a marker for the outer mitochondrial membrane⁷. Thus it seems that this protein is a constituent of the outer mitochondrial membrane. Little or no protein was found in the soluble fraction and it could not be solubilized unless a detergent (0.1% v/v Triton X-100) was present in the buffer.

Mitochondria from a wide variety of tissues can accumulate large amounts of calcium⁸⁻¹⁰ and they have long been thought to be involved in calcium transport across intestinal cells^{11,12}. However, the influence of vitamin D_3 and $1,25-(\text{OH})_2\text{D}_3$ on the

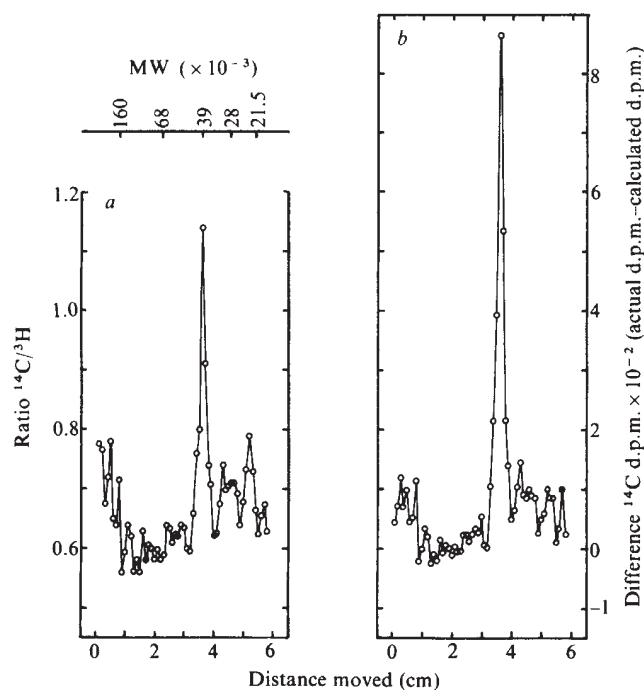


Fig. 1 Mitochondrial proteins synthesized in intestinal slices from $1,25-(\text{OH})_2\text{D}_3$ -dosed or deficient chickens. Flasks containing 0.5 g of everted chick jejunal slices (about 2 mm thick) were incubated in 10 ml of Krebs Improved Ringer 1 medium¹⁸ for 2 h with shaking in O_2/CO_2 (19:1) atmosphere at 37°C . Slices from vitamin D-deficient chicks or those dosed 4 h before death with 125 ng of $1,25-(\text{OH})_2\text{D}_3$ were incubated with $[4,5-^3\text{H}]\text{leucine}$ (125 μCi) and $[\text{U}-^{14}\text{C}]\text{leucine}$ (25 μCi), respectively. After incubation the slices were washed several times with 0.9% w/v NaCl and their mitochondria isolated¹⁹. Mitochondria from dosed and deficient chicks were mixed (50,000–100,000 d.p.m. of ^3H , 25,000–50,000 d.p.m. of ^{14}C and ~ 0.5 mg protein) and dissociated before electrophoresis by heating at 70°C for 30 min in a solution containing dithiothreitol (15 mg ml^{-1}), SDS (10 mg ml^{-1}), Tris (12 mg ml^{-1}) and glycerol (0.1 ml). After electrophoresis in SDS-polyacrylamide⁵ the gel was washed for several hours in acetic acid/propan-2-ol/water (2:5:13) by volume and cut into 1-mm sections. Each slice was treated with 0.5 ml 90% (v/v) NCS tissue solubilizer (Amersham/Searle) at 60°C for 2 h in closed vials. After cooling, 0.5 ml of NCS tissue solubilizer and 10 ml of scintillator containing 0.01% 1,4-bis-(5-phenyloxazol-2-yl) benzene and 0.4% 2,5-diphenyloxazole in toluene was added to the digested gel and radioactivity estimated in a Packard Tri-Carb scintillation spectrometer model 2650. Quenching corrections were made by means of automatic external standard and correlation curves for combined ^{14}C and ^3H . a, Ratio of $^{14}\text{C}:^3\text{H}$ in each slice of gel; b, difference between observed ^{14}C radioactivity in each slice and that expected if each slice had the average ratio of $^{14}\text{C}:^3\text{H}$ (ref. 6).

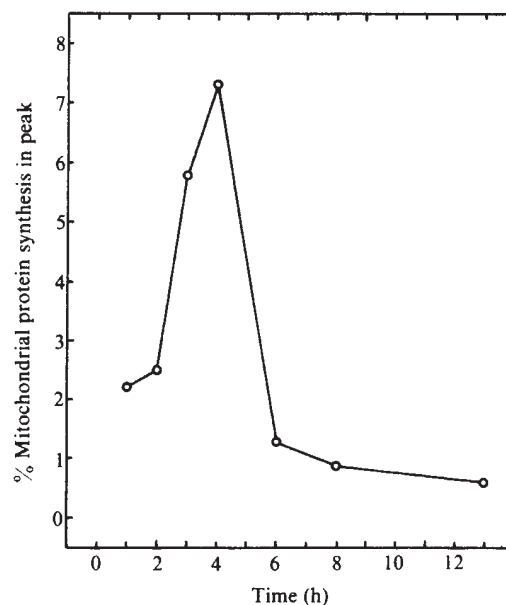


Fig. 2 Time course of synthesis of mitochondrial protein. Slices of everted chick jejunum from vitamin D-deficient chicks or those dosed at various times before death with 125 ng of $1,25-(\text{OH})_2\text{D}_3$ were incubated with radioactive leucine and their mitochondria prepared as described in Fig. 1 legend. Mitochondrial proteins from chicks dosed with $1,25-(\text{OH})_2\text{D}_3$ at a certain time before death were mixed with those from vitamin D-deficient chicks electrophoresed and counted as described previously (Fig. 1 legend). The level of ^{14}C radioactivity observed in the peak at MW 39,000–42,000 in a difference plot⁵ (see Fig. 1b) was estimated and expressed as percentage of total ^{14}C on the gel.

mitochondria of intestinal cells has received little attention. Vitamin D_3 administered *in vitro* and *in vivo* seems to alter the ability of kidney mitochondria to retain calcium *in vitro*^{13,14} and some authors report that CaBP can influence the release of calcium from intestinal mitochondria¹¹. In addition, a small yet direct effect of $1,25-(\text{OH})_2\text{D}_3$ on calcium uptake by intestinal mitochondria *in vitro* has been demonstrated¹⁵. In this experiment it was suggested that the direct effect might be of only secondary importance—the steroid caused a much greater increase in calcium uptake merely by increasing cytoplasmic concentrations of calcium. Such an effect would account for the

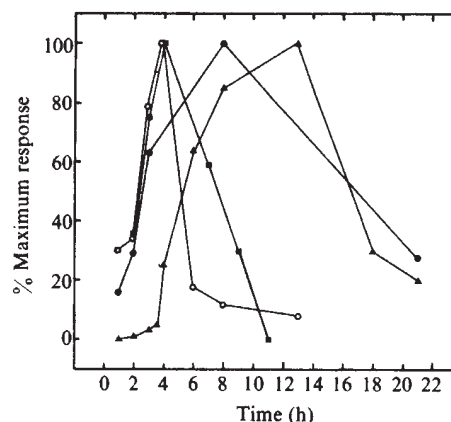


Fig. 3 Time course of events occurring in the rachitic chick intestine after a single intracardial injection of $1,25-(\text{OH})_2\text{D}_3$ (125 ng). $\circ$, Synthesis of mitochondrial protein; $\blacksquare$, synthesis of brush border protein⁴; $\bullet$, calcium transport *in vitro*³; $\blacktriangle$, synthesis of CaBP (ref. 3).

transient accumulation of calcium by intestinal mitochondria *in vivo* after administration of vitamin D₃ (ref. 16).

We have been unable to find any role for this novel mitochondrial protein in calcium transport. Although the protein may be required to regulate some aspect of mitochondrial calcium transport, most studies have suggested that these processes are controlled both by the electrochemical gradient which is established across the inner mitochondrial membrane and by a calcium carrier, associated with the inner membrane, which catalyses a continuous efflux of calcium (for a review see ref. 17). The outer mitochondrial membrane is considered to be of little importance in calcium transport. However, the evidence for CaBP stimulation of calcium release from mitochondria¹¹ suggests an intermediary role for the mitochondrial protein in this process. Alternatively it may act independently of CaBP and sequester calcium from the mitochondrial intramembrane space thus facilitating calcium efflux and calcium transport across the intestinal cell. We do not know whether the protein is produced *de novo* or merely at a greater rate after *in vivo* administration of 1,25-(OH)₂D₃. Further experiments are required to answer these questions.

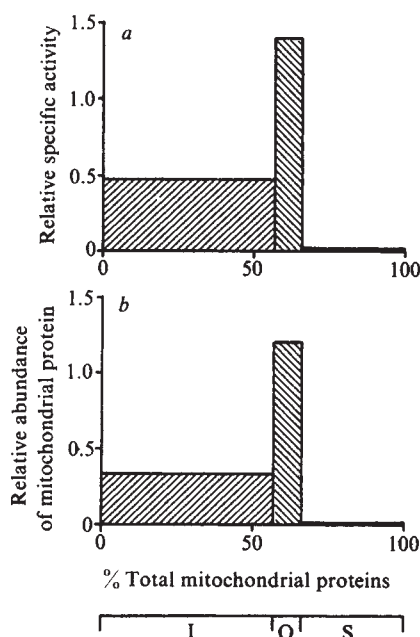


Fig. 4 Distribution of newly synthesized protein between sub-mitochondrial fractions. Mitochondria from the jejuna of vitamin D-deficient chicks ([4,5-³H]leucine-labelled) or from such chicks dosed 4 h before death with 125 ng 1,25-(OH)₂D₃ ([U-¹⁴C]leucine-labelled) as described in Fig. 1 legend were mixed 2:1 dosed: deficient (by weight) and sub-mitochondrial particles were prepared essentially by the method of Sottocasa *et al.*⁷. The mixture of mitochondria (20 mg protein ml⁻¹) was resuspended in 7.5 ml 10 mM Tris-phosphate pH 7.4 at 20 °C, left on ice for 10 min and mixed with 2.5 ml 1.8 M sucrose, 2 mM (Na)₂ATP, 2 mM MgSO₄ followed by a further 10 min on ice. Aliquots (3.5 ml) of this final mixture were sonicated on ice for three 15-s periods, with 30-s rests between each sonication, using a 9.5-mm diameter probe at an amplitude of 4 μm in a MSE ultrasonic disintegrator Mk 2. The sonicated mitochondria were then pooled and layered onto 8.5 ml 0.76 M sucrose, 5 mM Tris-phosphate pH 7.4, which was itself layered on 14 ml 1.32 M sucrose, 5 mM Tris-phosphate pH 7.4 in a Beckman SW 27 tube. After centrifugation for 3 h at 24,000 r.p.m. and 4 °C in a Beckman SW 27 rotor three fractions were collected—the pellet, designated the inner mitochondrial membrane fraction (I); material at the interface of the 0.76M and 1.32M sucrose, designated outer mitochondrial membrane fraction (O); and the top layer, designated soluble proteins (S). Portions of each fraction were assayed for: a, the presence of rotenone-insensitive NADH-cytochrome *c* reductase⁷ or b, the percentage of total ¹⁴C-labelled protein synthesis present in the peak at MW 39,000–42,000 following SDS-gel electrophoresis and a difference plot (see Figs 1, 2).

Received 21 August; accepted 27 October 1980.

1. Wasserman, R. H., Corradino, R. A. & Taylor, A. N. *J. biol. Chem.* **243**, 3978–3986 (1968).
2. Wasserman, R. H., Taylor, A. N. & Fullmer, C. S. *Biochem. Soc. spec. Publ.* **3**, 55–74 (1974).
3. Spencer, R., Charman, M., Wilson, P. & Lawson, E. *Nature* **263**, 161–163 (1976).
4. Wilson, P. W. & Lawson, D. E. M. *Biochim. biophys. Acta* **497**, 805–811 (1977).
5. Wilson, P. W. & Lawson, D. E. M. *Biochem. J.* **173**, 627–631 (1978).
6. Mayol, R. F. *Molec. cell. Endocr.* **2**, 133–146 (1975).
7. Sottocasa, G. L., Kuylenstierna, B., Ernster, L. & Bergstrand, A. *J. Cell Biol.* **32**, 415–438 (1967).
8. Lehninger, A. L., Carafoli, E. & Rossi, C. S. *Adv. Enzym.* **29**, 259–321 (1967).
9. Lehninger, A. L. *Biochem. J.* **119**, 129–138 (1970).
10. Bygrave, F. L. *Biol. Rev.* **53**, 43–79 (1978).
11. Hamilton, J. W. & Holdsworth, E. S. *Biochem. biophys. Res. Commun.* **40**, 1324–1330 (1970).
12. Omdahl, J. L. & De Luca, H. F. *Physiol. Rev.* **53**, 327–372 (1973).
13. Engstrom, G. W. & De Luca, H. F. *J. biol. Chem.* **237**, 974–975 (1962).
14. Engstrom, G. W. & De Luca, H. F. *Biochemistry* **3**, 203–209 (1964).
15. Bikle, D. D., Askew, E. W., Zolock, D. J., Morrissey, R. L. & Herman, R. H. *Biochim. biophys. Acta* **598**, 561–574 (1980).
16. Sampson, H. W., Matthews, J. L., Martin, J. H. & Kunin, A. S. *Calcified Tissue Res.* **5**, 305–316 (1970).
17. Nicholls, D. G. & Crompton, M. *FEBS Lett.* **111**, 261–268 (1980).
18. Dawson, R. M. C. in *Data for Biochemical Research* 2nd edn (eds Dawson, R. M. C., Elliot, C. D., Elliot, W. H. & Jones, K. M.) 476–508 (Clarendon, Oxford, 1969).
19. Freeman, K. B. *Biochem. J.* **94**, 494–501 (1965).

Role of *src* gene in growth regulation of Rous sarcoma virus-infected chicken embryo fibroblasts

Gordon Parry, James C. Bartholomew & Mina J. Bissell*

Laboratory of Chemical Biodynamics, Lawrence Berkeley Laboratory, University of California, Berkeley, California 94720

We report here a study of the mechanisms leading to loss of growth control in chicken embryo fibroblasts transformed by Rous sarcoma virus (RSV). We have been particularly concerned with the role of the *src* gene in this process, and have used RSV mutants temperature sensitive (*ts*) for transformation to investigate the nature of the growth regulatory lesion. The two principal findings were (1) the stationary phase of the cell cycle (G₁) in chick embryo fibroblasts seems to have two distinct regulatory compartments (using the terminology of Brooks *et al.*¹ we refer to these as 'Q' and 'A' states). When rendered stationary at 41.5 °C by serum deprivation, normal cells enter a Q state, but cells infected with the *ts*-mutant occupy an A state. (2) Whereas normal cells can occupy either state depending on culture conditions, the *ts*-infected cells, at 41.5 °C, do not seem to enter Q even though a known *src* gene product, a kinase, is reported to be inactive at this temperature^{2,3}. We discuss the possibility that viral factors other than the active *src* protein kinase influence growth control in infected cultures.

The experiments involved plating cultures at high density and low serum content in 35-mm plates in medium containing 0.5% chicken serum (see Fig. 1 legend), and maintaining them at either 41.5 or 35 °C for 36–48 h before starting an experiment. DNA synthesis was followed by staining cells with propidium iodide and analysing cellular fluorescence by flow cytometry. After 48 h in 0.5% serum, cells infected with the *ts* mutants LA24, LA29 or NY68 mainly occupied the G₁ compartment if cultured at 41.5 °C. However, a significant proportion was found in S if cultures were maintained at 35 °C (Table 1). Cultures maintained at 41.5 °C and then shifted to 35 °C were found to enter S approximately 8 h after shift (Fig. 1). These data confirm the early work of Bell *et al.*⁴. The lag period before entry into S was distinctly longer when normal cells seeded at high density

* To whom reprint requests should be addressed.

Table 1 Growth regulation in ts-mutants LA29 and LA24 (Prague A) and NY68 (Schmidt Ruppin A)

	Fraction of the population in cell cycle phase		
	G ₁	S	G ₂ + M*
41.5 °C			
Normal cells	0.694	0.027	0.279
NY 68	0.688	0.028	0.304
LA 29	0.688	0.042	0.271
LA 24	0.688	0.028	0.304
35 °C			
Normal cells	0.703	0.012	0.286
NY 68	0.423	0.324	0.253
LA 29	0.407	0.478	0.114
LA 24	0.498	0.424	0.078

Chicken embryo fibroblasts were cultured as described in Fig. 1 legend except that 4 h after plating, some cultures were shifted to 35 °C whereas others were maintained at 41.5 °C. Cultures were kept at these temperatures for 48 h and the distribution of cells throughout the cycle analysed after this time period, using the procedures described in Fig. 1 legend. The cells in the G₂ + M category include some G₁ doublets which were not dissociated in our experimental conditions and thus the values in this category do not reflect true G₂ + M estimations.

* Doublet population significant.

and maintained in low serum were stimulated to initiate DNA synthesis by serum addition at 41.5 °C. In six separate experiments, normal cells at 41.5 °C were found to enter S phase 12–14 h after stimulation, but surprisingly, LA24-infected cells entered S phase only 6–8 h after stimulation (Fig. 2). These results implied that in the stationary state normal and LA24-infected cultures were at different stages of the G₁ phase.

If normal cells were initially plated at a lower density or serum stimulated 24 h after plating instead of 48 h, then their lag period before entry into S approached that of LA24-infected cultures (Fig. 3). This clearly demonstrated that there are two

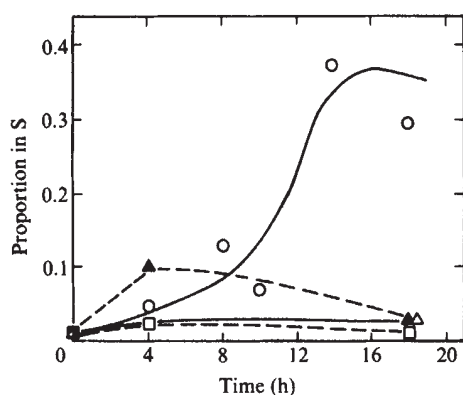


Fig. 1 Kinetics of entry of chicken embryo fibroblasts into S phase after shift from 41.5 to 35 °C. Tertiary cells were plated in medium 199 supplemented with 2% tryptose broth, 0.5% chicken serum and 0.1% glucose, at a density of 2×10^6 per 35-mm plate. They were kept at 41.5 °C for 36–38 h and either maintained at this temperature for a further 18 h or shifted to 35 °C. At given times, plates were removed and the cells collected and stained with propidium iodide as described elsewhere¹¹. The stained cells were subsequently analysed in a flow cytometer and histograms of the distribution of cells throughout the cell cycle obtained. The proportion of cells in S was determined using a fitting procedure, and data analysed using a program described elsewhere¹². ○, LA24-infected cells cultured at 41.5 °C and shifted to 35 °C; △, normal cells cultured at 41.5 °C and shifted to 35 °C; □, LA24-infected cells cultured at 41.5 °C and held at that temperature, and ▲, normal cells cultured at 41.5 °C and held at that temperature.

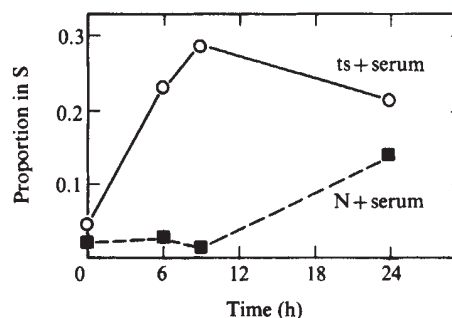


Fig. 2 Serum stimulation of DNA synthesis in normal and LA24-infected cells maintained at 41.5 °C. Cells were seeded at 2.0 – 2.5×10^6 per 35-mm dish, cultures as in Fig. 1 and maintained at 41.5 °C for 48 h. DNA synthesis was initiated by the addition of 200 µl of calf serum to the cultures (no medium change: 2 ml medium per plate) and the proportion of cells in S at times after addition was determined as described in Fig. 1 legend. ○, LA24-infected cells; ■, normal cells (N).

compartments in the stationary phase of the chick embryo fibroblast cell cycle and that, by manipulating culture conditions, normal cells can be made to occupy either compartment. In contrast, it was not possible to force the LA24-infected cells into the compartment with a 12–14 h lag even by plating the cells at much higher densities than the normal cells or by growing them strictly at 42 °C. Thus, viral factors in the stationary ts-infected cells prevent these cells from entering the state which would necessitate the longer lag period. Experiments with cells infected with transformation-defective virus showed that their lag period from serum stimulation to entry into S was similar to that of normal cells (data not shown).

That chicken embryo fibroblasts apparently have two distinct stationary states within the G₁ phase is consistent with the recent work of Brooks *et al.*¹, who discuss the evidence for two stationary states (Q and A) in mammalian cell-cycle models. Our data suggest that this is also true for the avian cell cycle, the Q state being a long-lag compartment and the A state the short-lag compartment.

Our data also indicate that virally coded factor(s) active at the non-permissive temperature must be responsible for maintaining the infected cells in the short-lag stationary stage in G₁ or preventing the cells from moving into the long-lag compartment. At present, the only known activity coded for by the *src* gene is a protein kinase, pp60^{src}, which has been shown to be temperature sensitive in ts-mutants³. It is possible to argue that a low level of kinase is expressed at 41.5 °C and that this is sufficient to advance the cells further along the cycle. Such levels, however, would have to be very small in LA24-infected cells because the *in vivo* (cellular) activity of the kinase has been shown to be similar to that in uninfected controls even at 41 °C. Furthermore, in experiments carried out with cells cultured at strictly 42 °C, when such 'leakiness' would be minimized, the same differences in cycle position were observed. Also, note that in earlier studies which examined the expression of many transformation parameters at permissive and non-permissive temperatures using LA24, we failed to detect leakiness of other usual transformation parameters at the non-permissive temperature^{6,7}.

A more feasible alternative is that an activity other than the kinase is responsible for these observations—this could be another activity associated with pp60^{src} at the non-permissive temperature or a different protein coded for by a different reading frame of the *src* gene. Support for this suggestion comes from the observation in neuroretinal cells that the growth-regulating activity of *src* and its other transformation-related

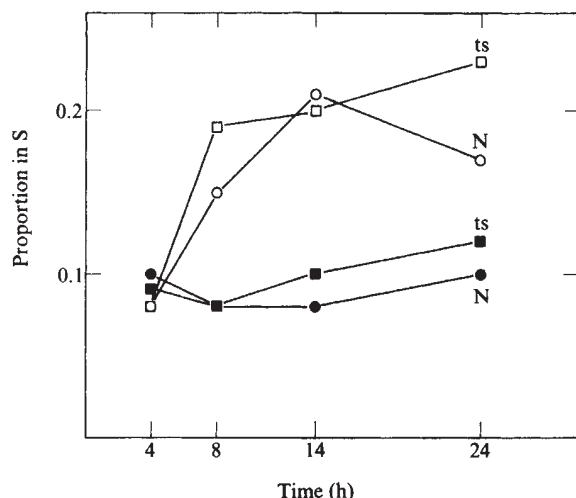


Fig. 3 Stimulation of DNA synthesis in normal and LA24-infected cells maintained at lower density at 41.5 °C. Cells were seeded at 1.2×10^6 per 35-mm plate in the conditions described in Fig. 1 legend. They were kept at 41.5 °C for 18–24 h. DNA synthesis was initiated by changing the medium with fresh medium containing 2–5% chick or calf serum. Proportion of cells in S was determined as described in Fig. 1 legend. Closed symbols, control cultures; open symbols, serum-stimulated cultures.

functions may be dissociable⁸. Calothy *et al.*⁸ have isolated an RSV mutant, PA-101, which is defective for 'transformation', but causes proliferation of neuroretinal cells. The 60,000 molecular weight (60K) *src* protein produced by these cells is inactive as a kinase in the immune complex assay (J. M. Bishop, personal communication).

Other observations on cells infected with temperature-sensitive RSV mutants are interesting in terms of these results. Poste and Flood⁹ found that infected cells formed tumours at both 41 and 35 °C when grown on chicken chorioallantoic membranes. Harley and Goldfine¹⁰ found that in contrast to several types of normal cells, infected cells need not synthesize lipids before initiating DNA synthesis when shifted from 41 to 35 °C. Finally, it was found that infected cells cultured at 41 °C were more sensitive to tumour promoters than were normal cells⁷. Data reported here, together with these observations, are consistent with the idea that cells infected with temperature-sensitive RSV mutants are in an initiated or committed state at the non-permissive temperature, and provide strong evidence for the multi-stage model of viral oncogenesis which has been discussed in detail elsewhere⁷.

We thank Jill Hatie and Deborah Farson for help with cell culture, Krishnakali Das for data analysis, and J. Michael Bishop for allowing us to cite unpublished data. This work was supported by the Division of Biomedical and Environmental Research, United States Department of Energy, under contract W-740-ENG-48, and an SRC (UK) fellowship to G.P.

Received 18 July; accepted 27 October 1980.

- Brooks, R. F., Bennett, D. C. & Smith, J. A. *Cell* **19**, 493–504 (1980).
- Collet, M. S. & Erikson, R. L. *Proc. natn. Acad. Sci. U.S.A.* **75**, 2021–2024 (1978).
- Rübsamen, H., Friis, R. R. & Bauer, H. *Proc. natn. Acad. Sci. U.S.A.* **76**, 967–971 (1979).
- Bell, J. G., Wyke, J. A. & Macpherson, I. A. *J. gen. Virol.* **27**, 127–134 (1975).
- Sefton, B. M., Hunter, T., Beemon, K. & Eckhart, W. *Cell* (in the press).
- Parry, G. & Hawkes, S. P. *Proc. natn. Acad. Sci. U.S.A.* **75**, 3703–3707 (1978).
- Bissell, M. J., Hatie, C. & Calvin, M. *Proc. natn. Acad. Sci. U.S.A.* **76**, 348–352 (1979).
- Calothy, G., Doirier, F., Dambrine, G. & Pessac, B. *Virology* **89**, 75–84 (1978).
- Poste, G. & Flood, M. K. *Cell* **17**, 789–800 (1979).
- Harley, J. B. & Goldfine, H. *Exp. Cell Res.* **118**, 47–54 (1979).
- Teng, M., Bartholomew, J. C. & Bissell, M. J. *Nature* **268**, 739–741 (1977).
- Bartholomew, J. C., Pearlman, A. L., Landolph, J. R. & Straub, K. *Cancer Res.* **39**, 2538–2543 (1979).

An embryo protein induced by SV40 virus transformation of mouse cells

Peter T. Mora, K. Chandrasekaran
& Vivian W. McFarland

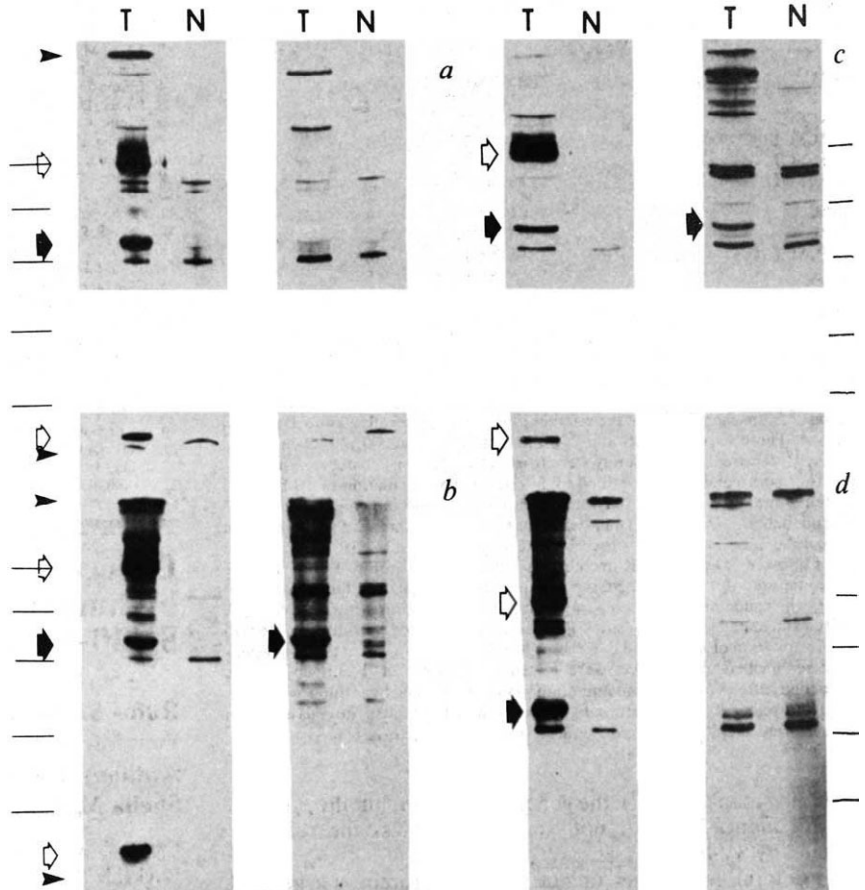
National Institutes of Health, National Cancer Institute, Bethesda, Maryland 20205

A specific protein of molecular weight (MW) ~55,000 (55K) was found recently by immunoprecipitation in all SV40 virus-transformed mammalian cells^{1–8}, in addition to the SV40 large T antigen (~94K) and small t antigen (~17K), which are the only proteins coded by the 'early half' of the SV40 genome. The 55K protein is encoded by cellular DNA; its peptide pattern is different from that of the SV40 antigens and it is species specific in mouse, rat, hamster, monkey and human SV40-transformed (or infected) cells^{6–9}. A 55K protein with a similar peptide pattern was found in mouse embryonal carcinoma cells not exposed to SV40^{4,5}. Similar proteins were reported in mouse sarcomas and leukaemias induced by a great variety of aetiological agents and also in a spontaneously transformed mouse fibroblast cell line¹⁰, and it has been suggested that the protein may be a general correlate of cellular tumorigenicity^{11–13}. We now report that the ~55K protein is present in primary cell cultures from 12–14-day old mouse embryos, but not in 16-day old mouse embryos. The embryo protein has a peptide pattern virtually indistinguishable from that of the SV40-induced protein. We also show by comparing closely related cell families that spontaneously transformed highly tumorigenic mouse cells do not possess the 55K protein.

The large T antigen, the ~55K cellular protein and the small t antigen were the major labelled proteins specifically immunoprecipitated from an SV40-transformed cloned mouse fibroblast cell *215CSC with the SV40 tumour (T) serum^{1–9}, as shown in the left half of the panels of Fig. 1. In the right half of Fig. 1a, the 'parent' clone 210C, not transformed by SV40, does not show these specifically immunoprecipitable proteins, and thus also lacks the 55K protein¹. The T serum used is able to precipitate the cellular 55K protein independently of the presence of SV40 large T antigen, as shown on the embryonal carcinoma cell line F9 (right half of Fig. 1b)^{4,5,14}.

Mouse embryos at various post-implantation stages were investigated and compared. Primary cultures from 12-day old mouse embryos showed the presence of specifically immunoprecipitable ~55K protein (see dark arrow in the right half of Fig. 1c). Similarly, 10–14-day mouse embryo primary cultures also possessed the 55K protein (data not shown). Equally or more highly radioactive immunoprecipitates from primary cultures from 16-day old mouse embryos exhibited no specific immunoprecipitable protein in the 55K region (right half of Fig. 1d). Apparently, beyond a certain embryonic developmental stage, significant amounts of the protein are no longer synthesized in primary cultures. When the primary embryo cultures from 12-, 14- or 17-day embryos were replated and labelled, none of the secondary or subsequent passages tested showed detectable significant amounts of 55K protein (data not shown; see also data on 210C and other established mouse and other embryo cell lines^{1–8}). Thus, the cells which synthesize detectable 55K cellular protein in primary culture either are no longer present in further cultures in sufficient amounts or do not synthesize the 55K protein in significant amounts. In preliminary experiments with M. Dziadek we have now detected the 55K protein in 13-day mouse embryo organ cultures of liver, brain and carcass, but not in similar 16-day old embryo organ cultures, indicating that the protein indeed correlates with stages

Fig. 1 Detection of the 55K protein in various cells. Preconfluent, exponentially growing cells in tissue culture²¹ were labelled for 3 h with L-[³⁵S]methionine, extracted, immunoprecipitated with either serum from hamster bearing SV40-induced tumour (T serum) or with normal (N) preimmune hamster serum, and subjected to SDS-polyacrylamide gel electrophoresis and fluorography, as described elsewhere¹. Left to right, pairwise: panel a, the SV40-transformed clonal mouse embryo cell line *215CSC and the parent clone 210C; panel b, *215CSC and F9 teratocarcinoma cells; panel c, *215CSC and 12-day old AL/N mouse embryo primary culture; panel d, *215CSC and 16-day old embryo primary (see Table 1 legend for the abbreviations used for the cells). Dark wide arrows indicate the ~55K cellular protein, the open arrows the position of the SV40 large T and small t antigens, and dark narrow arrows the top and bottom of the separating gels. Lines on the side of the panels show the positions of the MW marker proteins: Phosphorylase a, 94K; bovine serum albumin, 66.5K; ovalbumin, 45K; DNase I, 31K; trypsinogen, 24K; lactoglobulin, 18K. Embryo cells were obtained as follows. A virgin 8-week old AL/N female was placed in a cage with one 8-week old AL/N male mouse; 16 h later the female was checked for vaginal plug and removed from the cage. This day was designated day 0. After various (10–17) days of pregnancy, the females were killed, the uterus excised and individual embryos (average seven or eight per mouse) removed in sterile conditions. Embryos were dissected free of extraembryonic membranes, combined, minced, washed with phosphate-buffered saline pH 7.4, gently trypsinized and plated for primary cultures, as described elsewhere²¹. In a procedure similar to that used for all the other cells after one change of growth medium, preconfluent exponentially growing cells of the primary cultures were labelled as above. Equal volumes of the cell extract supernatants were immunoprecipitated with T or N serum, as described previously¹. The input radioactivities of the T serum precipitates were 30–40% higher than the N serum precipitates, resulting in the more intense protein bands in the former. Overexposed fluorograms (7 days) are presented to facilitate detection of any (possible) 55K proteins in the N serum precipitates.



of embryogenesis and that its synthesis in the appropriate primary cultures is not a result of culture artefacts (for example, cell 'stimulation' or selection).

To obtain enough labelled 55K embryo protein for a proteolytic peptide comparison, it was necessary to have at least three synchronously pregnant AL/N mice and combine 12-day old embryos for primary culture. The 55K band, separated on an SDS-polyacrylamide gel as in Fig. 1, was cut out and eluted from the gel. A similar amount of labelled SV40-induced protein was obtained from *215CSC cells. The two labelled proteins were subjected separately to limited proteolysis with increasing amounts of *Staphylococcus aureus* V8 protease and re-electrophoresed¹⁵. Figure 2 shows seven to nine methionine-labelled peptides with identical mobilities, appearing symmetrically with increased enzyme concentrations, from the

SV40-transformed *215CSC cells (left half) and the 12-day old embryo primary cells (right half of fluorogram). Thus, the 55K protein in 12-day old embryos is very similar (or identical) to the SV40-induced (or stabilized) 55K cellular protein. Preliminary findings with D. Simmons also show very similar (or identical) two-dimensional electrophoresis and chromatography patterns⁸ for the methionine-labelled tryptic peptides of the two proteins. Comparison of the proteins from SV40-transformed cells from different species suggests that the protein is an evolutionarily conserved protein⁸, and might thus have an essential cellular function. The results indicate that the host gene(s) for such protein(s) is functioning constitutively in the embryo during a certain phase of embryogenesis. In established embryo cells

Fig. 2 Comparison of peptides obtained by partial proteolysis of the isolated 55K proteins from 12-day old embryo primary cells and from the SV40-transformed *215CSC cells. Twenty 12-day old AL/N embryos were used for primary cultures and labelled as in Fig. 1. The *215CSC cells were also labelled similarly. Immunoprecipitates with T serum were separated as in Fig. 1. The 55K protein bands were located, cut out, homogenized in a buffer containing 10 mM NaHCO₃, 0.1% SDS, 0.1% β -mercaptoethanol and 20% sucrose¹⁵, and extracted at 37 °C overnight. The extraction mixture was centrifuged and the supernatant sampled for counting. The total radioactivities recovered were 35,889 c.p.m. from the embryo cells and 23,940 c.p.m. from the *215CSC cells. Four aliquots (20 μ l) each of the two supernatants, containing 8,545 c.p.m. (embryo) and 5,700 c.p.m. (*215CSC), were placed in wells in the 3% spacing gel. Homogenizing buffer¹ (10 μ l) containing 0, 45, 90 and 450 ng of *Staphylococcus aureus* V8 protease (Miles) was placed in the wells as shown on the top of the figure. The wells were filled with a sample electrophoresis buffer²², subjected to electrophoresis (30 V, 3 h) until the marker dye had penetrated 1 inch of the stacking gel. Electrophoresis continued at 140 V for 2 h, until the dye reached the bottom of the gel. The gel was stained, destained and fluorographed as before¹. Exposure was for 20 days. The solid arrow points to the undigested 55K protein present in largest amount in the gel columns from the buffer controls.

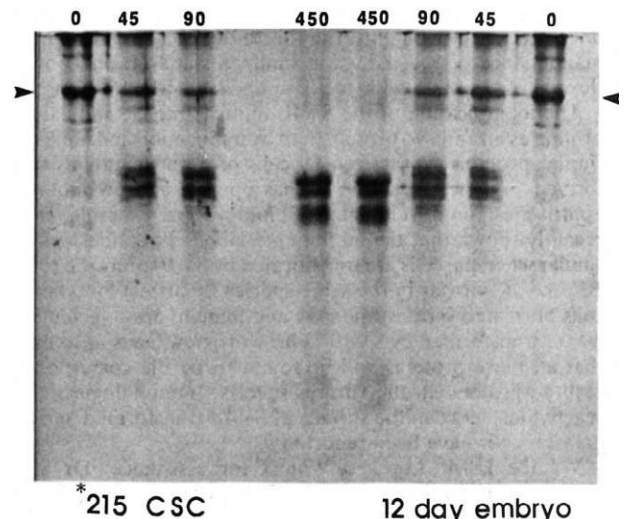


Table 1 Absence of correlation between tumorigenicity and the 55K protein

Cell	TD ₅₀	55K protein	Cell	TD ₅₀	55K protein
104C	2	—	*215CSC	6.4	+
*106CSC	4	+	*221CSCT	2	+
124CSCT	2	—	*222CSCT	<3	+
127CSCT	<2	—	*223CSCT	3	—
128CSCT	<2	—	*224CSCT	2	+
134CSCTC	>4	—	219CT	<2	—
210C	6.4	—	301T	4	—
213CSC	5.5	—	*303TCS	>5	+
*214CSC	>6	+	*314TSCT	>5	+
			*315TSCT	5	+

Properties of mouse embryo cell lines from three cell 'families'. The relation and generation of the cells are indicated with the postfix letters in sequence: C denotes a clone (obtained by a single-cell technique²⁰); S, SV40 transformed; C, recloned; T, tumour cell line obtained from the clone immediately above (by injecting 10⁵–10⁷ cells into a syngeneic mouse and re-establishing a cell line from the tumour¹⁹). An asterisk before the number denotes that the cell is SV40 T-antigen positive. The cells in the 100 and 200 families were from AL/N strain mouse embryo¹⁹; those in the 300 family were from BALB/c mouse embryo. The TD₅₀ denotes tumorigenicity, determined by intramuscular inoculation of 10-fold cell dilutions (minimum of three each) into batches of 6–8-week old syngeneic mice (10 mice per batch): TD₅₀ = log₁₀[number of cells producing rapidly lethal sarcomas in 50% of the mice inoculated]; observed and calculated as in ref. 19. The presence (+) or absence (–) of the 55K protein was determined as in Fig. 1. The SV40 T antiserum used (lot 78 × 222, NCI) was capable of precipitating the 55K protein alone, when such protein was separated by gel electrophoresis from the SV40 T antigen and eluted separately (as in Fig. 2) and reprecipitated (not shown) (see also the precipitation of the F9 and embryo cells in Fig. 1). Also, when the T-antigen-negative labelled cell extracts were mixed with excess of T-antigen-positive unlabelled cell extracts no more immunoprecipitable radioactive counts were seen in the 55K region than without mixing (data not shown). Some of the cells had been characterized previously for biochemical, tissue culture and *in vivo* growth properties^{19,20}.

(including many clones¹), the genes are present, but the proteins are not induced and/or not 'stabilized', unless the cells are infected by SV40.

To determine whether or not the 55K protein is a general correlate of cellular tumorigenicity, the presence of the 55K protein in three families of established mouse fibroblast cells were compared with tumorigenicity in the syngeneic mouse. There was no correlation at all in the mouse embryo fibroblast cells tested (Table 1). The highly malignant sarcomagenic cells in all three families from two strains (AL/N and BALB/c) of mice, which became tumorigenic by spontaneous transformation (cells 104C, 124CSCT, 127CSCT, 128CSCT, 219CT and 301T), did not possess detectable amounts of 55K protein when tested with the SV40 T antiserum, whereas the correlation was found without exception in all three families with the cells expressing SV40 T antigen and the 55K protein¹. Tumour cell lines 124, 127 and 128CSCT and the 134CSCTC clone carry late SV40 DNA, but are revertant in the sense that they do not possess the SV40 early gene¹⁶. This shows that the continued presence of at least one copy¹⁶ of the SV40 early gene (as in *106CSC) and its protein product (T antigen) is sufficient and, in established cell lines, necessary not only for the initiation⁵ but also for the maintenance of the induction (and/or stabilization—possibly through interaction with the T antigen¹⁷) of the 55K protein.

The identification of the SV40-induced mouse cellular 55K protein as an embryo protein (and by implication and analogy to similar proteins synthesized by cells of other mammals transformed by other means^{6–11,13}) has opened up new molecular approaches for developmental biology. For example, it was recently shown that the methionine-labelled peptides are very similar when the cells are transformed by SV40 or by the related BK and JC viruses in the same species (hamster), or when the cells of related species (monkey and human) are infected lytically or transformed by SV40⁹. Our work now leads us to expect that all these proteins will be found to be the corresponding embryo proteins in the various species. Immunological cross-reactivities between the surface of SV40-transformed and fetal hamster cells have been reported¹⁸.

We thank Mr Lorenzo Waters for assistance, Dr S. W. Luborsky for discussion, and Drs D. Simmons and M. Dziadek for allowing us to quote results.

Received 18 June; accepted 6 November 1980.

- Chang, C., Simmons, D. T., Martin, M. A. & Mora, P. T. *J. Virol.* **31**, 463–471 (1979).
- Kress, M., May, E., Cassingena, R. & May, P. J. *J. Virol.* **31**, 472–483 (1979).
- Melero, J. A., Stitt, D. T., Mangel, W. F. & Carroll, R. B. *Virology* **93**, 466–480 (1979).
- Linzer, D. I. H. & Levine, A. J. *Cell* **17**, 43–52 (1979).
- Linzer, D. I. H., Maltzman, W. & Levine, A. J. *Virology* **98**, 308–318 (1979).
- Smith, A. E., Smith R. & Paucha, E. *Cell* **18**, 335–346 (1979).
- Melero, J. A., Tur, S. & Carroll, R. B. *Proc. natn. Acad. Sci. U.S.A.* **77**, 97–101 (1980).
- Simmons, D. T., Martin, M. A., Mora, P. T. & Chang, C. J. *J. Virol.* **44**, 650–657 (1980).
- Simmons, D. J. *J. Virol.* **36**, 519–525 (1980).
- DeLeo, A. J. *et al. Proc. natn. Acad. Sci. U.S.A.* **76**, 2420–2424 (1979).
- Jay, G. *et al. Cold Spring Harb. Symp. quant. Biol.* **44**, 469–664 (1980).
- Pollack, R. *et al. Cold Spring Harb. Symp. quant. Biol.* **44**, 681–688 (1980).
- Luka, J., Jörnvall, H. & Klein, G. J. *J. Virol.* **35**, 592–602 (1980).
- Segal, S., Levine, A. J. & Khoury, G. *Nature* **280**, 335–338 (1979).
- Cleveland, D. W., Fischer, S. G., Kirschner, M. W. & Laemmli, U. K. *J. biol. Chem.* **252**, 1102–1106 (1977).
- Küster, J. M., Mora, P. T., Brown, M. & Khoury, G. *Proc. natn. Acad. Sci. U.S.A.* **74**, 4796–4800 (1977).
- Lane, D. P. & Crawford, L. V. *Nature* **278**, 261–263 (1979).
- Weppner, W. A. & Coggin, J. H. *Jr Cancer Res. 40*, 1380–1387 (1980).
- Mora, P. T., Chang, C., Couvillion, L., Küster, J. M. & McFarland, V. W. *Nature* **269**, 36–40 (1977).
- Winterbourne, D. J. & Mora, P. T. *J. biol. Chem.* **253**, 5109–5120 (1978).
- McFarland, V. W., Mora, P. T., Schultz, A. & Pancake, S. J. *cell. Physiol.* **85**, 101–111 (1975).
- Laemmli, U. K. *Nature* **227**, 680–685 (1970).

Defective repair of alkylated DNA by human tumour and SV40-transformed human cell strains

Rufus S. Day III*, Chuck H. J. Ziolkowski*, Dominic A. Scudiero†, Sharon A. Meyer†, Anthony S. Lubiniecki‡, Anthony J. Girardi§, Sheila M. Galloway|| & Gaither D. Bynum¶

* Nucleic Acids Section, LMC, CIP, DCCP, NCI, NIH, 9000 Rockville Pike, Bethesda, Maryland 20205

† Chemical Carcinogenesis Program, Frederick Cancer Research Center, Frederick, Maryland 21701

‡ Flow Laboratories, 7655 Old Spring House Road, McLean, Virginia 22101

§ Institute for Medical Research, Copewood and Davis Streets, Camden, New Jersey 08103

|| Litton Bionetics, 5516 Nicholson Lane, Kensington, Maryland 20795

¶ Section on Cellular Aging, Laboratory of Comparative Cellular Physiology, Gerontology Research Center, NIA, Baltimore, Maryland 21224

We have identified a group of 8 (among 39) human tumour cell strains deficient in the ability to support the growth of adenovirus 5 preparations treated with N-methyl-N'-nitro-N-nitrosoguanidine (MNNG), but able to support the growth of non-treated adenovirus normally^{1,2}. This deficient behaviour defines the Mer⁻ phenotype². Strains having the Mer⁻ phenotype were found to arise from tumours originating in four different organs. Relative to Mer⁺ strains, Mer⁻ tumour strains showed greater sensitivity to MNNG-produced killing, greater MNNG-stimulated 'DNA repair' synthesis and a more rapid MNNG-produced decrease in semi-conservative DNA synthesis². Here we report that (1) Mer⁻ strains are deficient in removing O⁶-methylguanine (O⁶-MeG) from their DNA after [Me-¹⁴C]MNNG treatment (Table 1); (2) Mer⁻ tumour strains originate from tumours arising in patients having Mer⁺ normal fibroblasts (Fig. 1a, b); (3) SV40 transformation of (Mer⁺) human fibroblasts often converts them to Mer⁻ strains (Fig. 1c, d); (4) MNNG produces more sister chromatid exchanges (SCEs) in Mer⁻ than in Mer⁺ cell strains (Fig. 2).

Mer⁺ and Mer⁻ strains were treated with ¹⁴C-MNNG and analysed directly following and 20-h after treatment for DNA-bound O⁶-MeG and N⁷-MeG (N⁷-methylguanine; see Table 1). O⁶-MeG is not removed appreciably from the DNA of Mer⁻ strains over the 20-h incubation period. However, the amount of O⁶-MeG (relative to N⁷-MeG) found in the DNA of MNNG-treated Mer⁺ strains shortly after MNNG treatment is about

threefold less than that found in Mer⁻ strains, presumably reflecting active removal by Mer⁺ strains of O⁶-MeG during the 1-h MNNG treatment (ref. 3 and B. Strauss, personal communication). During the 20-h incubation after MNNG treatment, the Mer⁺ strains continued to remove O⁶-MeG from their DNA. By contrast, in the Mer⁻ strains, the O⁶-MeG/N⁷-MeG ratio increased, presumably reflecting the slow loss of N⁷-MeG from the DNA expected at 37°C (refs 3-6). (Our data are consistent with a half life for DNA-bound N⁷-MeG of ~46 h, which is in the range of those published^{5,6}.)

We questioned whether the tumours that gave rise to the Mer⁻ tumour strains arose in persons who were entirely composed of Mer⁻ cells (indicating possible genetic inheritance of the condition) or whether the non-tumour cells of such patients would be Mer⁺ (indicating that the Mer⁻ condition may be a property of only tumour cells). To investigate this question experimentally, we obtained strains of normal fibroblasts prepared from two patients whose tumours had given rise to the Mer⁻ A427N and Hs750T strains. The normally appearing fibroblast strains, A427F and Hs750M, were both found to be Mer⁺ (Fig. 1a). Studies of post-MNNG colony-forming ability (Fig. 1b) established that the fibroblast lines A427F and Hs750M had sensitivities to MNNG similar to those of human fibroblasts (see refs 2, 7, and Fig. 1d). The A427N tumour line, however, showed the increased sensitivity to MNNG-produced

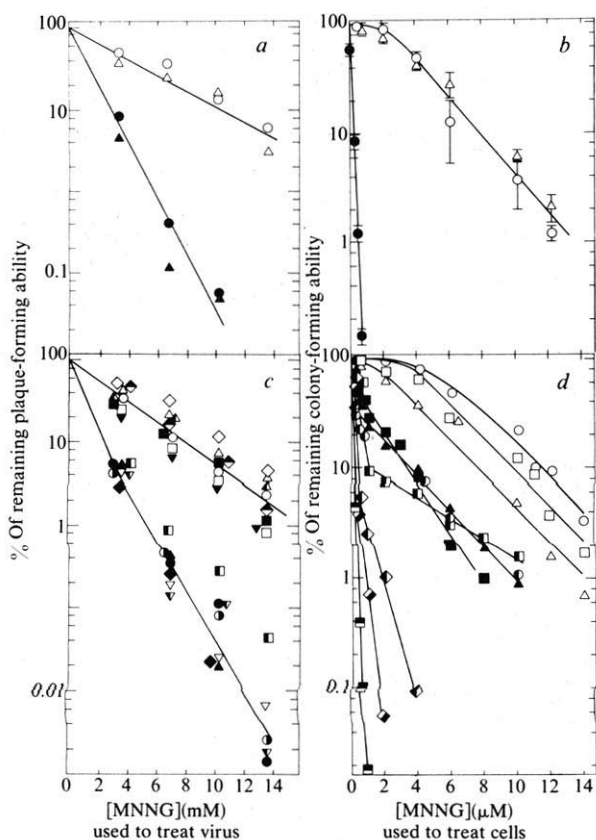


Fig. 1 a, b, MNNG-produced inactivation of adenovirus 5 (a) and of colony-forming ability (b) using normal or tumour cell strains from the same patient. Patient 1: ○, A427F (normal fibroblast strain) and ●, A427N (lung tumour strain). Patient 2: △, Hs750M (normal fibroblast strain) and ▲, Hs750T (stomach tumour strain). For a, adenovirus 5 was inactivated with MNNG and subsequently assayed for plaque-forming ability using monolayers of the above strains as described¹⁹. For colony-forming ability (b), cells were seeded at low density, treated with MNNG and grown to colonies as described elsewhere⁷. No colony data are presented for Hs750T because it did not form colonies. Virus plaquing efficiencies and cellular colony-forming efficiencies were in the ranges of those reported previously and were not correlated with the sensitivities observed. c, d, MNNG-produced inactivation of adenovirus 5 (c) and of colony-forming ability (d) using normal human fibroblasts, SV40-transformed fibroblasts and Mer⁻ tumour strains. The cell strains used were: ◇, AT5BIVA2; △, WI38; ▲, WI38VA13; ◆, GM54; ■, GM638; ○, IMR90; ● and ○, IMR90-830; ●, IMR90-890; □, WI26; ■, WI26VA4; ▲, GM637; ▽, XP12; ▽, W98VA1; ◆, SV80; ▽, WI8VA2; ■, A172; ◆, A875; ◆, BE.

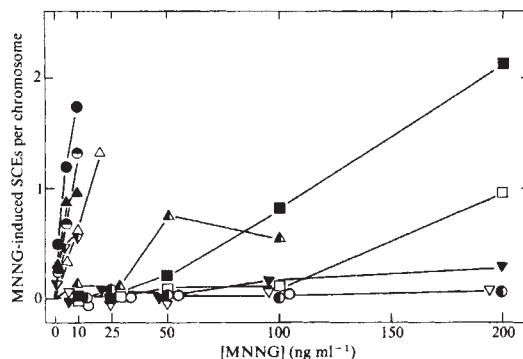


Fig. 2 Number of MNNG-induced SCEs as a function of MNNG concentration used to treat cells of five Mer⁻ and seven Mer⁺ strains. Confluent cells were split into 75-cm² flasks containing DMEM or McCoy's 5A plus 10% fetal calf serum. The next morning, stocks of MNNG were diluted into the medium to obtain the desired concentrations. Bromodeoxyuridine was added to 10 μg ml⁻¹. After 57-80 h of incubation, colcemid (10 μg ml⁻¹, Gibco) was added to 0.1-0.2 μg ml⁻¹. After 4 h further incubation, the cells were collected, centrifuged and resuspended in 5 ml 0.04-0.075 M KCl, centrifuged and fixed in 5 ml methanol/glacial acetic acid, 3:1, and dried on to slides. For differential staining, slides were either immersed in 4'-6-diamidino-2-phenylindole at 10 μg ml⁻¹ of distilled water for at least 3 min, removed and treated with three drops 0.2 M NaPO₄ buffer, pH 11.5, or stained by a modified Hoechst 33258-Giemsa technique^{20,21}. The chromosomes of at least 15 metaphase cells were counted to determine each point. Standard deviations were typically 0.2-0.3 of the average. The cell strains used, followed by the zero dose level of SCEs per chromosome, were: Mer⁻: ●, A1336 (0.34); ●, A2095 (0.14); ▽, A1235 (0.18); ▲, WI38VA13 (0.38); △, A253 (0.26); Mer⁺: □, A388 (0.35); ○, IMR90 (0.16); ■, A549 (0.15); ●, A673 (0.14); ▲, A704 (0.20); ▽, HuTu80 (0.16); and ▽, W98VA1 (0.64).

killing characteristic of Mer⁻ tumour strains (ref. 2 and Fig. 1d); tumour strain Hs750T did not form colonies. Therefore, the respective patients were not entirely composed of Mer⁻ cells, showing that generalized inheritance of the Mer⁻ condition did not occur.

SV40-transformed human fibroblast strains were tested for their Mer phenotype to determine whether the Mer⁻ phenotype was characteristic only of certain strains derived from human tumours, or if human strains transformed by other means were also Mer⁻. Of 11 SV40-transformed strains, 7 were found to be Mer⁻ and 4 Mer⁺ (Fig. 1c Table 1). The available non-transformed human fibroblast 'parents' of these transformed strains (Table 2) were all Mer⁺, including three strains (IMR90, WI38 and GM54) that gave rise to Mer⁻ transformed strains (Table 2). The result indicates that SV40 transformation often converts Mer⁺ strains to Mer⁻ strains, or selects Mer⁻ variants as targets for transformation.

The SV40-produced Mer⁻ strains were not as sensitive as Mer⁻ tumour strains to the killing effects of MNNG (Fig. 1d), suggesting that either the Mer⁻ tumour strains are defective in the ability to remove an additional MNNG-produced lesion, or that the cellular response of the two groups of Mer⁻ strains to O⁶-MeG differs. Further, although the SV40-produced Mer⁻ cells were more MNNG sensitive than their parent fibroblasts, they were not more sensitive to MNNG-produced killing than were Mer⁺ SV40-transformed cells, and, therefore, were not killed by O⁶-MeG, which they are deficient in excising. (To be certain that the Mer⁻ phenotype of the tumour strains was not caused by adventitious SV40 infection, we found, using our assay⁸, that SV40 T-antigen was not detectable in any of the eight Mer⁻ tumour strains described above. By contrast, and as a control, more than 95% of the nuclei of the SV80 strain contained T-antigen.)

Sister chromatid exchanges^{9,10} reflect cellular DNA damage due to various mutagenic and carcinogenic agents^{10,11} and are increased in damaged cells lacking certain DNA repair processes^{12,13}. Figure 2 shows that Mer⁻ strains, including one (WI38VA13) derived from normal fibroblasts (WI38), gave rise to detectable levels of SCEs at much (up to 250-fold) lower concentrations of MNNG than did Mer⁺ strains. However, there

Table 1 O^6 - and N^7 -methylguanines in MNNG-treated Mer⁺ and Mer⁻ human cell strains

	After 1 h MNNG treatment (c.p.m.)			After 1 h MNNG treatment + 20 h incubation (c.p.m.)		
	N^7 -MeG	O^6 -MeG	O^6 -MeG/ N^7 -MeG	N^7 -MeG	O^6 -MeG	O^6 -MeG/ N^7 -MeG
Mer⁻ strains						
A427N	1,221.5	139.5	0.114	1,164.1	169.1	0.145
A253	1,365.0	132.8	0.097	963.8	129.0	0.134
SV80	419.0	33.1	0.079	319.3	36.5	0.114
A172	643.3	43.9	0.068	398.0	57.9	0.145
HeLa S3	933.8	84.7	0.091	493.4	51.2	0.104
A2095	271.6	30.6	0.113	145.8	19.7	0.135
A875	712.3	65.5	0.092	272.6	37.0	0.136
A1617	734.3	72.1	0.098	440.2	51.4	0.117
BE	762.7	74.1	0.097	400.2	62.8	0.157
GM638	1,182.9	129.1	0.109	845.9	126.4	0.149
IMR90-830	1,334.0	156.1	0.117	521.0	67.1	0.129
W18VA2	1,664.2	186.2	0.112	1,199.4	163.0	0.136
W138VA13	1,183.1	105.6	0.089	760.9	99.6	0.131
Average			0.098			0.133
Mer⁺ strains						
HeLa CCL2	926.2	25.2	0.027	529.2	10.0	0.019
A549	1,007.2	39.7	0.039	832.9	12.0	0.014
HT29	1,519.5	26.2	0.017	927.3	8.1	0.009
H4	1,020.8	35.3	0.035	613.6	10.7	0.017
HuTu80	388.0	6.3	0.017	170.9	2.6	0.015
W98VA1	3,183.3	133.7	0.042	2,004.8	21.9	0.011
A388	514.4	27.0	0.052	434.6	8.4	0.019
A704	328.2	8.7	0.027	175.2	0.9	0.005
Average			0.032			0.014

Amounts of O^6 - and N^7 -methylguanines in the DNA of cell strains treated for 1 h with 6 μ M [Me - ^{14}C]MNNG. Confluent cells ($1-5 \times 10^6$ on 18–20 100-mm culture dishes per strain) were treated for 1 h with 6 μ M [Me - ^{14}C]MNNG, 15 Ci mol⁻¹, NEN 100 μ Ci ml⁻¹ ethanol) in a solution containing 50% phosphate-buffered saline (PBS) and 50% Dulbecco's modified minimal essential medium (MEM, the latter containing 10% fetal calf serum, 50 μ g streptomycin per ml, 50 units penicillin per ml and 40 μ g gentamycin per ml, DMEM). Half of the plates were then refed with DMEM and incubated overnight at 37 °C. The cells (1 h) from the remaining plates were washed once with 5 ml PBS, then taken up with trypsin-EDTA, centrifuged, resuspended in 10 ml PBS, centrifuged again, resuspended in 5 ml 0.01 M Tris-0.02 M EDTA, pH 8.0 (TE) and frozen. About 20 h later, the incubated treated cells were collected similarly but not frozen. The first sample of cells was thawed and DNA was extracted from both samples as follows: 0.5 ml TE containing 2% SDS, 0.5 ml TE containing 500 μ g RNase A (Sigma) and 0.5 ml TE containing 1,200 units T₁ RNase (Sigma) were added to each sample; this was mixed and incubated for 1 h at 37 °C. Proteinase K (1 mg, Merck) in 1 ml TE was then added, followed by 1 h incubation at 37 °C. Phenol (redistilled, saturated with 0.25 M Tris-HCl, pH 8.0, 5 ml) was added to each sample, which was then shaken for 10 min at room temperature. NaCl (5 M, 1.5 ml) was added to the separated aqueous layer, followed by mixing, the addition of 14 ml ice-cold ethanol and winding out the DNA. The DNA was resuspended in TE overnight, precipitated with 5% trichloroacetic acid or ethanol and treated with 60 μ l 0.1 M HCl at 70 °C for 30 min to remove the purines. Any sediment that formed after 30 min in ice was removed by centrifugation. To the supernatant was added 12 μ l 1 M NaH₂PO₄ and 12 μ l of 100 μ g ml⁻¹ O^6 -MeG (marker). The purines were separated by passage through a Waters μ Bondapak C18 column (0.39 $\times$ 30 cm) using a Waters liquid chromatography apparatus (model U6K injector, model 440 absorbance detector used at 254 nm, model 6000A pump, model 730 data module) with 0.01 M NH₄H₂PO₄ plus 16% (final) methanol as solvent at 1 ml min⁻¹. In these conditions, N^3 -methyladenine eluted at about 4.5 min, N^7 -MeG at 7 min and O^6 -MeG at 11.5 min. Fractions (0.5 or 1.0 ml) were collected, and where needed, brought to a volume of 1.0 ml with solvent and counted (Beckman LS7500) in 10 ml Aquasol (NEN) at 61% efficiency. Background (8.7–9.0 c.p.m.) is subtracted. The initial reaction (1 h) converted $0.5-2.0 \times 10^{-5}$ of the total guanine to O^6 -MeG (Mer⁻) and $0.8-2.0 \times 10^{-6}$ of the guanine to N^7 -MeG.

Table 2 Human strains used in the present study

Tumour strains			Fibroblast strains		
Mer ⁻	Tumour of origin	Source	Mer ⁺	Tumour of origin	Source
A427N	Lung carcinoma	a	Hs750M	Muscle tissue	a
A253	Epidermoid carcinoma, neck	a	A427F	Lung	b
A172	Astrocytoma	a	WI26	Embryonic lung	e
HeLa S3	Cervical carcinoma	c	WI38	Embryonic lung	e
A2095	Synovial cell sarcoma	b	IMR90	Embryonic lung	f,d
A875	Melanoma	b	GM54	Human	f
A1617	Adenocarcinoma, colon	b	SV40-transformed strains		
BE	Colon carcinoma	d			
Hs750T	Stomach carcinoma, metastatic to lung	a	Mer ⁻	Fibroblast of origin	
A1336	Ovarian carcinoma	b	SV80	Fetal fibroblast strain	a
A1235	Astrocytoma	b	XP12	XP12RO	c
Mer ⁺			GM638	GM54	f
HeLa CCL2	Cervical carcinoma	c	IMR90-830	IMR90	g
A549	Lung carcinoma	b	IMR90-890	IMR90	g
HT29	Colon carcinoma	d	W138VA13	WI38	e,g
H4	Neuroglioma	a	W18VA2	W18	g
HuTu80	Stomach tumour	a	Mer ⁺		
A388	Epidermoid carcinoma	a,b	AT5BIVA2	AT5BI	g
A704	Kidney carcinoma	a,b	W98VA1	W98	g
A673	Rhabdomyosarcoma	a	W126VA4	W126	g
			GM637	GM37	c,f

a, Supplied by Drs Walter Nelson-Rees and Jack Weaver, with support from the NCI Viral Oncology Program, under the auspices of the Office of Naval Research and the Regents of the University of California. b, Gifts of Dr S. A. Aaronson, NCI. c, Gifts of Dr R. M. Baker, Center for Cancer Research, MIT, Cambridge, Massachusetts. d, Gifts of Dr Kurt Kohn, NCI. e, Obtained from the American Type Culture Collection, Maryland. f, Obtained from the Institute of Medical Research, Camden, New Jersey. g, Prepared by A.G. using procedures similar to those published¹⁸. The last eight Mer⁻ tumour strains are reported here for the first time as having the Mer⁻ phenotype. This fact was established for each line in at least two experiments of the type shown in Fig. 1a, c.

is great heterogeneity in response among the Mer⁺ strains, indicating that factors other than the DNA repair mechanism governed by the Mer system influence the yield of induced SCEs due to any one dose of MNNG.

Thus, we have shown here that the Mer⁻ phenotype is distinguished from Mer⁺ by deficient removal of O^6 -MeG produced by MNNG in their DNA, and by increased suscep-

tibility to MNNG-produced SCEs in addition to the differences noted in our introduction. Further, the demonstration of the Mer⁻ behaviour of the HeLa S3 and XP12 strains, known to be more mutable by alkylating agents than either the HeLa CCL2 or the GM637 strain^{14,15}, provides a link between the Mer⁻ phenotype and increased sensitivity to mutagenesis by alkylating agents.

The two SV40-transformed human strains, identified here as Mer⁻ (XP12, derived from a xeroderma pigmentosum (XP) patient) and Mer⁺ (GM637, derived from a control person), are divergent both in their susceptibilities to MNNG-produced SCEs (XP12 was the more sensitive¹²) and in their abilities to remove O⁶-alkylguanines from their DNA^{4,16} (GM637 showed such excision, XP12 did not). These differences were provisionally hypothesized to be another manifestation of the repair defect in XP fibroblasts that leads to deficient repair of UV ray-damaged DNA. We believe that these differences are more probably due to the SV40-produced molecular changes that lead to the expression of the Mer⁻ phenotype.

The identification and characterization of the Mer⁻ phenotype in many transformed human cell lines, but not in normal human fibroblasts, leave us unable to decide whether the Mer⁻ repair deficiency enhances (or is produced by) transformation, or whether virally produced transformation often leaves a Mer⁻ 'footprint'. That 1-chloroethyl-1-nitrosourea (CNU) produces

more DNA cross-links in Mer⁻ than in Mer⁺ tumour strains (see accompanying manuscript¹⁷) further suggests that the presence or absence of the Mer repair system may be the molecular reason for the *in vivo* failure or success of chemotherapeutic regimens using agents (1,3-bis-chloroethylnitrosourea (BCNU) or 3-cyclohexyl-1-chloroethyl-1-nitrosourea (CCNU)) having modes of action similar to that of CNU.

We thank Drs Kurt Kohn, R. M. Baker and Len Erickson for helpful discussions. Dr Kohn first suggested investigating the Mer phenotype of WI38VA13 and Dr Baker provided strains HeLa CCL2, HeLa S3, GM637, and XP12 and data before publication. We also thank Dr S. A. Aaronson for the supply of human tumour cell strains, all of which begin with 'A' in this work, Dr Hermann Autrup both for providing us with N³-MeA, N⁷-MeG and O⁶-MeG standards and for encouragement, Dr Tsuyoshi Kakefuda for instruction in HPLC techniques and for his patience, and Dr Ken Kraemer for suggestions concerning the manuscript.

Received 29 July; accepted 27 October 1980.

- Day, R. S. III & Ziolkowski, C. H. J. *Nature* **279**, 797-799 (1979).
- Day, R. S. III, Ziolkowski, C. H. J., Scudiero, D. A., Meyer, S. A. & Mattern, M. R. *Carcinogenesis* **1**, 21-32 (1980).
- Montesano, R., Bresil, H. & Margison, G. P. *Cancer Res.* **39**, 1798-1802 (1979).
- Bodell, W. J., Singer, B., Thomas, G. H. & Cleaver, J. E. *Nucleic Acids Res.* **6**, 2819-2829 (1979).
- Altamirano-Dimas, M., Sklar, R. & Strauss, B. *Mutat. Res.* **60**, 197-206 (1979).
- Singer, B. J. *natn. Cancer Inst.* **62**, 1329-1339 (1979).
- Scudiero, D. A. *Cancer Res.* **40**, 984-990 (1980).
- Kaplan, M. M., Giard, D. J., Blattner, W. A., Lubiniecki, A. S. & Fraumeni, J. F. Jr *Proc Soc. exp. Biol. Med.* **148**, 660-664 (1975).
- Latt, S. A. *Proc. natn. Acad. Sci. U.S.A.* **71**, 3162-3166 (1974).
- Perry, P. & Evans, H. J. *Nature* **258**, 121-125 (1975).

- Castrano, A. V., Thompson, L. H., Lindl, P. A. & Minkler, J. L. *Nature* **270**, 551-553 (1978).
- Wolff, S. A., Rodin, B. & Cleaver, J. E. *Nature* **265**, 347-349 (1977).
- De Weerd-Kastelein, E. A., Keijzer, W., Rainaldi, G. & Bootsma, D. *Mutat. Res.* **45**, 253-261 (1977).
- Baker, R. M., Van Voorhis, W. C. & Spencer, L. A. *Proc. natn. Acad. Sci. U.S.A.* **76**, 5249-5253 (1979).
- Baker, R. M., Zuerndorfer, G. & Mandel, R. *Environ. Mutagenesis* **2**, 269-270 (1980).
- Goth-Goldstein, R. *Nature* **267**, 81-82 (1977).
- Erickson, L. C., Laurent, G., Sharkey, N. A. & Kohn, K. *Nature* **288**, 727-729 (1980).
- Girardi, A. J., Jensen, F. C. & Koprowski, H. *J. cell. comp. Physiol.* **65**, 69-84 (1965).
- Day, R. S. III *Photochem. Photobiol.* **19**, 9-13 (1974).
- Goto, K., Maeda, S., Kano, Y. & Sugiyama, T. *Chromosoma* **66**, 351-359 (1978).
- Perry, P. & Wolff, S. *Nature* **251**, 156-158 (1974).

DNA cross-linking and monoadduct repair in nitrosourea-treated human tumour cells

Leonard C. Erickson*, Guy Laurent, Nancy A. Sharkey & Kurt W. Kohn

Laboratory of Molecular Pharmacology, Developmental Therapeutics Program, Division of Cancer Treatment, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20205

The 1-(2-chloroethyl)-1-nitrosoureas are potent anti-cancer drugs^{1,2} which produce DNA inter-strand cross-links in a two-step reaction sequence. The first step was proposed to be an addition of a chloroethyl group to a guanine-O⁶ position of DNA³; the second step, which occurs over a period of several hours in the absence of free drug³⁻⁶, could then form an inter-strand cross-link by the slow reaction of the bound chloroethyl group with a nucleophilic site on the opposite DNA strand. The delay between the formation of chloroethyl monoadducts and the formation of inter-strand cross-links allows time for a DNA repair mechanism, capable of removing the monoadducts, to prevent the cross-linking. We recently proposed this mechanism to account for a difference in inter-strand cross-linking between a normal and a transformed human cell strain⁶. Day and his coworkers (see refs 7, 8 and previous paper⁹) found that some human tumour cell strains (designated Mer⁻ phenotype) are deficient in the ability to repair O⁶-methylguanine lesions in DNA. We therefore hypothesized that the repair function that removes O⁶-methylguanine residues from DNA would also remove chloroethyl monoadducts and hence prevent chloroethylnitrosourea-induced inter-strand cross-linking. We now present evidence that supports this hypothesis and indicates also that the O⁶-methylguanine repair confers resistance to cell killing by chloroethylnitrosourea.

Thirteen human cell strains, most of them derived from malignant tumours, were studied (Table 1). The phenotype of each strain was classified as Mer⁺ or Mer⁻ by R. Day, on the

basis of whether or not the cells were proficient in supporting the growth of adenovirus 5 treated with N-methyl-N'-nitro-N-nitrosoguanidine⁷⁻⁹.

The cells were treated with 1-(2-chloroethyl)-1-nitrosourea (CNU) for 1 h and then incubated for 6 h to allow inter-strand cross-links to form^{5,6}. Inter-strand cross-linking and DNA-protein cross-linking were determined by the alkaline elution method¹⁰⁻¹², using the same procedures as in our previous study⁶. The elution assay, when performed directly (without proteinase), measures the combined effect of inter-strand and DNA-protein cross-links. The major contribution to this measurement, in the case of chloroethylnitrosoureas, is from DNA-protein cross-links^{5,6}. The assay with proteinase greatly reduces or eliminates the DNA-protein cross-links, and is assumed to be a measure of inter-strand cross-linking¹¹.

The results show a clear correlation between inter-strand

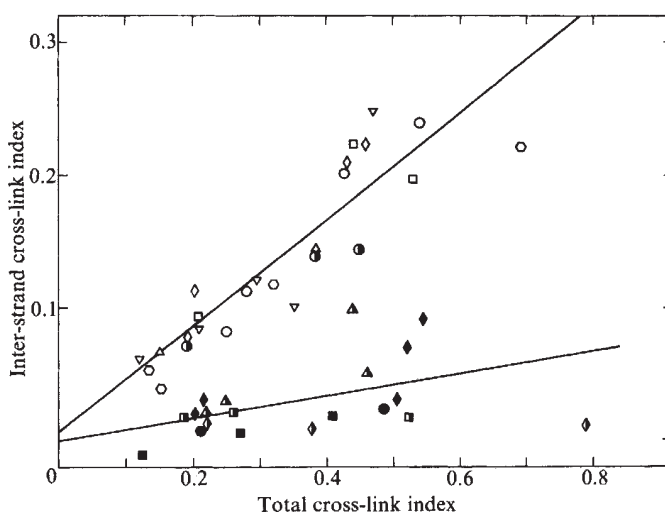


Fig. 1 Inter-strand cross-linking (assay with proteinase) plotted against DNA-protein cross-linking, estimated by assay without proteinase in the same experiment. Cross-link index is defined in Table 1 legend. Open symbols, Mer⁻ strains; closed and half-closed symbols, Mer⁺ strains. The symbols representing individual cell strains are shown in Table 1.

* To whom correspondence should be addressed at: Building 37, Room 5D17, National Institutes of Health, National Cancer Institute, Bethesda, Maryland 20205.

Table 1 Alkaline elution assays of DNA inter-strand and DNA-protein cross-linking by 1-(2-chloroethyl)-1-nitrosourea (CNU) (cross-link index $\times 10^3$)

Cell strain	Origin	Mer phenotype	Inter-strand cross-linking (proteinase assay)		Total cross-linking (no proteinase)	
			50 μ M CNU	100 μ M CNU	50 μ M CNU	100 μ M CNU
■ HT29	Colon ca.	+	0 (−13–+13) ³	7 (0–17) ³	126, 135	268, 410
● IMR90	Embryonic lung	+	3 (−3–+7) ³	23 (23–24) ³	210, 216	485
◇ A2182	Lung ca.	+	14 (8–20) ³	31	154, 218	788
▣ A673	Rhabdomyosarcoma	+	16 (12–20) ³	16	186, 261	523
◆ A549	Lung ca.	+	33 (20–54) ⁴	52 (19–91) ⁴	203, 216	504, 545
△ HT1080	Fibrosarcoma	+	62 (20–107) ⁴	127 (51–230) ³	221, 251	437, 462
○ A431	Epidermoid ca. (vulva)	+	69 (60–76) ³	143 (142–144) ²	190, 190	385, 446
○ A1336	Ovarian ca.	−	69 (38–177) ³	221	135, 152	692
▽ A427	Lung ca.	−	72 (61–84) ⁴	149 (101–247) ⁴	122, 208	295, 471
□ BE	Colon ca.	−	81 (71–93) ³	201 (182–224) ⁴	166, 207	438, 530
◇ A172	Astrocytoma (grade IV)	−	84 (62–113) ³	217 (209–225) ²	193, 204	432, 457
△ VA13	SV40-transformed WI-88	−	88 (50–146) ³	232 (145–320) ²	148, 152	384
○ A875	Melanoma	−	112 (82–142) ³	240 (201–281) ³	250, 281	540

Cells previously labelled with [2-¹⁴C]thymidine (0.01 μ Ci ml^{−1}, 24 h) were treated with 50 or 100 μ M CNU for 1 h in culture medium (Dulbecco's modified Eagle's medium, 10% heat-inactivated fetal calf serum, 0.02 M HEPES buffer) and then incubated for 6 h in drug-free medium. Control cells were similarly run without drug treatment. After collection, 2.5×10^5 cells were mixed with 5×10^5 L1210 mouse leukaemia cells whose DNA had been labelled with ³H-thymidine (0.1 μ Ci ml^{−1}, 24 h); these cells served as internal standards. The mixed cell suspensions were exposed to 300 R of ¹³⁷Cs γ -ray irradiation at 0 °C just before alkaline elution, which was carried out by the same procedures as used in our previous study⁶. Cells were deposited on 2- μ m-pore size polyvinyl chloride filters and lysed with 2% SDS, 0.02 M EDTA, 0.1 M glycine, pH 10, with or without 0.5 mg ml^{−1} proteinase K. Alkaline elution was with tetrapropylammonium hydroxide/0.02 M H₄EDTA, pH 12.1, pumped at a rate of 2 ml h^{−1}. When proteinase K was used, the eluting solution also contained 0.1% SDS. When proteinase K was not used, the lysis solution was washed out with 0.04 M Na-EDTA, pH 10, before elution at pH 12.1. Cross-link index is defined as $[(1 - R_0)/(1 - R_1)]^{1/2} - 1$, where R_0 and R_1 are the 'relative retention' for untreated and CNU-treated cells, respectively⁵. Relative retention is the fraction of the ¹⁴C-DNA remaining on the filter when 30% of the internal standard ³H-DNA remained on the filter. The range of values obtained in independent experiments are given in parentheses; the superscript indicates the number of experiments. Symbols in the first column on the left are used to represent the cell strains in Figs 1 and 2. ca., carcinoma. Cell strains: IMR90, normal human embryo cells obtained from W. Nichols; VA13, an SV40-transformed derivative of normal human embryo strain WI38; BE, obtained from B. Giovannella; HT29, obtained from E. Jensen. The other strains were obtained from R. Day.

cross-linking (proteinase assay) and methylation repair capacity (Table 1). The six strains deficient in this capacity (Mer[−] phenotype) produced consistently higher inter-strand cross-linking levels than did five of the seven Mer⁺ strains studied.

The remaining two Mer⁺ strains (A431 and HT1080) gave deviant results. Strain A431 yielded high inter-strand cross-linking values, in the range of the Mer[−] strains. In view of this result, Day (personal communication) restudied the A431 culture, and found that with continued passage the culture had attained a large Mer[−] component. In the case of the HT1080 strain, the first two inter-strand cross-link determinations were low, well within the Mer⁺ range, whereas the last two determinations were high, well within the Mer[−] range. This observation raises the question of whether cells in culture can lose the Mer⁺ phenotype.

In addition to assays for inter-strand cross-links, most of the experiments included assays for DNA-protein cross-links. The assays for total cross-links (no proteinase), which we take to be essentially measures of DNA-protein cross-linking, showed no significant differences between the Mer⁺ and Mer[−] strains (Table 1). Hence, the differences in inter-strand cross-linking cannot be attributed simply to differences in drug uptake or inactivation.

DNA-protein cross-links form without delay^{5,6} and probably involve different sites of linking to DNA from inter-strand cross-links. If the differences in inter-strand cross-linking are principally due to differences in a specific DNA repair mechanism, DNA-protein cross-linking, apparently little affected by such repair, could be a measure of the extent of drug entry and intracellular reactivity in particular experiments. The use of DNA-protein cross-linking as a gauge of intracellular drug delivery is plausible also because chloroethylnitrosoureas are lipid soluble, and hence their entry into cells is not likely to be limited by membrane barriers.

Using this concept, the value for inter-strand cross-linking (proteinase assay) was plotted against the value for DNA-protein cross-linking (no proteinase) obtained in the same experiment (Fig. 1). The cell strains are seen to give a bimodal distribution, the Mer[−] strains (open symbols) being clearly separated from most of the Mer⁺ strains (closed and half-closed symbols). The A431 (Mer⁺) strain again falls within the Mer[−] group (half closed circles). The segregation of the cell strains into two distinct groups is analogous to the distinct grouping with respect to methylation repair noted by Day and his

coworkers^{7–9}. This lends additional support to the idea that the cells differ with respect to a discrete phenotype.

The cytotoxic sensitivity of the cell strains to CNU was determined in terms of inhibition of cell proliferation (Fig. 2). Results with this assay have been found to be consistent with colony formation assays⁶. The six Mer[−] strains (open symbols) were clearly more sensitive to CNU than were five of the seven Mer⁺ strains.

The exceptions again were strains A431 (Fig. 2, half-closed circles) and HT1080 (Fig. 2, half-closed triangles). The A431 strain exhibited low sensitivity at low drug concentrations and high sensitivity at high concentration. It is possible that this strain has a limited capacity for O⁶-alkyl guanine repair, so that at low levels of DNA damage it seems to be Mer⁺, but at high levels seems to be Mer[−].

Our results, together with those of Day and his colleagues^{7–9}, suggest that the loss of a specific DNA repair function could simultaneously make cells susceptible to neoplastic transformation and make the resulting tumours vulnerable to

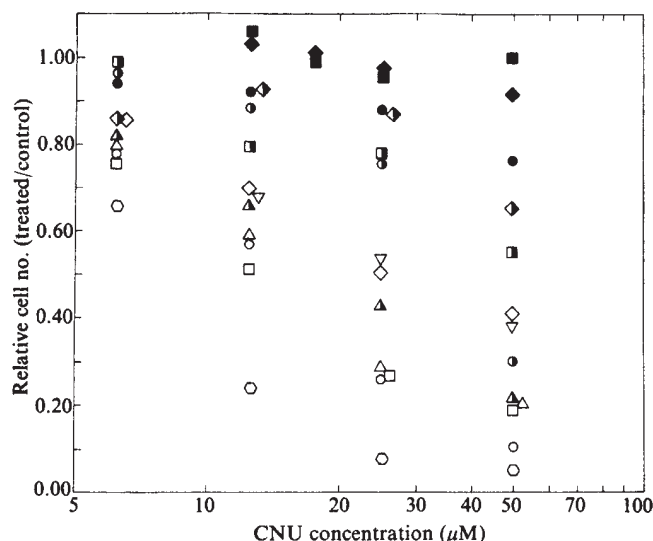


Fig. 2 Effect of CNU treatment on proliferation of various cell strains. Cells were treated with CNU for 1 h. After drug removal, the cells were allowed to proliferate in fresh medium for 3 days (three to five doubling times for control cells). Symbols are as in Fig. 1 and Table 1.

chemotherapy with certain DNA damaging agents. The loss of the ability to repair O^6 -alkyl guanine lesions in DNA could make the cells susceptible to the mutagenic effects of methylating agents and related carcinogens. Such cells would have an increased likelihood of malignant transformation, and could give rise to the Mer⁻ tumours. Our results suggest that such tumours may be vulnerable to DNA cross-linking by 1-(2-chloroethyl)-1-nitrosoureas.

Recent evidence indicates that repair of O^6 -methylguanine and O^6 -ethylguanine involves destruction rather than excision of these residues from DNA¹³⁻¹⁵. As O^6 -alkylation markedly alters the bond conjugation in the guanine ring, this could be the recognition feature for an enzyme which acts on the ring regardless of the identity of the alkylating group. Thus, O^6 -methylguanine and O^6 -2-chloroethylguanine residues might be destroyed by the same enzyme.

We thank Dr Rufus Day for providing cell strains which he had determined to have the Mer⁺ or Mer⁻ phenotype, and for many fruitful discussions. G.L. was supported by grants from the Fogarty International Center (NIH) and the Rose and Jean Hoguet Foundation.

Received 15 July; accepted 27 October 1980.

1. Wheeler, G. P. *Handbk exp. Pharmac.* **38**, 65-84 (1975).
2. Carter, S. K. *Cancer Treat. Rep.* **60**, 645-811 (1976).
3. Kohn, K. W. *Cancer Res.* **37**, 1450-1454 (1977).
4. Lown, J. W., McLaughlin, L. W. & Chang, Y.-M. *Bioorgan. Chem.* **7**, 97-110 (1978).
5. Ewig, R. A. G. & Kohn, K. W. *Cancer Res.* **38**, 3197-3203 (1978).
6. Erickson, L. C., Bradley, M. O., Ducore, J. M., Ewig, R. A. G. & Kohn, K. W. *Proc. natn. Acad. Sci. U.S.A.* **77**, 467-471 (1980).
7. Day, R. S. III & Ziolkowski, C. H. *Nature* **279**, 797-799 (1979).
8. Day, R. S. III, Ziolkowski, C. H., Scudiero, D. A., Meyer, S. A. & Mattern, M. R. *Carcinogenesis* **1**, 21 (1980).
9. Day, R. S. III *et al.* *Nature* **288**, 724-729 (1980).
10. Kohn, K. W., Erickson, L. C., Ewig, R. A. G. & Friedman, C. A. *Biochemistry* **15**, 4629-4637 (1976).
11. Kohn, K. W. *Meth. Cancer Res.* **16**, 291-345 (1979).
12. Kohn, K. W., Ewig, R. A. G., Erickson, L. C. & Zwelling, L. A. in *A Laboratory Manual of Research Procedures* (eds Friedberg, E. C. & Hanawalt, P. C.) (Dekker, New York, in the press).
13. Robins, P. & Cairns, J. *Nature* **280**, 74-76 (1979).
14. Karran, P. & Lindahl, T. *Nature* **280**, 76-77 (1979).
15. Renard, A. & Verly, W. G. *FEBS Lett.* **114**, 98-102 (1980).

Transient haem-globin interactions in photodeligated carboxyhaemoglobin and subunits

L. Lindqvist*, S. El Mohsni†, F. Tfibel* & B. Alpert†

* Laboratoire de Photophysique Moléculaire du CNRS, Université de Paris-Sud, 91405 Orsay Cédex, France

† Laboratoire de Biologie Physico-Chimique, UER Biomédicale, 45 rue des Saints-Pères, 75270 Paris Cédex 06, France

Although the fixation of ligand to haemoglobin (Hb) is known to be accompanied by changes in protein conformation¹ regulating the oxygen exchange in blood, the mechanism triggering these changes remains undecided²⁻⁴. We now report a dynamic approach to this problem using results obtained in a nanosecond laser photolysis study of carboxyhaemoglobin (HbCO) and its isolated subunits. The study is based on our previous observation⁵ of a structural evolution of free Hb after photodeligation, manifested through slight variations of the protein spectrum in the microsecond range. It is now found that the isolated subunits also show this behaviour. The duration of the spectral evolution is $\sim 2\mu\text{s}$ for the three proteins and the activation energy of the process $\sim 9\text{ kcal mol}^{-1}$. The spectral evolution is attributed to local conformation changes at the haem region, occurring during the structural relaxation of the freshly deligated protein. The results for the isolated chains show that such changes exist even in the absence of the R-T transition.

Hb and its isolated subunits were prepared from fresh human adult blood using standard procedures^{6,7}. Aqueous solutions of the proteins at pH 7.0 (0.1 M potassium phosphate buffer) were degassed by pumping, and potassium dithionite was added to

eliminate the last traces of oxygen. The CO-complexes were obtained by exposure of these solutions to CO; dissolved excess CO was removed by rough pumping, and an atmosphere of argon was applied above the solutions. The experiments were performed at 22 °C except where otherwise stated.

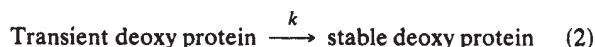
Photolysis light pulses of wavelength 530 nm and pulsewidth 7 ns at half-maximum were obtained from an Nd-glass laser. Maximum energy incident in the samples was 100 mJ. Optical transmission changes in the irradiated solutions (across 1 mm) were measured as a function of time at different wavelengths⁵.

We previously⁵ showed that photodeligated Hb initially has an absorption spectrum slightly different from that of stable deoxy Hb (Fig. 1a); however, this difference disappears within a few microseconds. We have established⁸ that this behaviour is not a multiphoton effect; such effects can always be suspected in pulsed laser studies⁹. The existence of a microsecond evolution of photodeligated Hb has also recently been confirmed by resonance Raman spectroscopy¹⁰.

Figure 1b and c show spectra measured after 100% photodeligation of the subunit CO complexes by laser excitation. The Soret band of the α chain is reduced in intensity compared with that of the stable deoxy protein, although to a smaller extent (7%) than that observed for Hb (18%). The corresponding Soret bands of the β chain seem to be coincident with each other. No distortion of the α and β subunit spectra was detectable in the 500-600 nm range, in contrast with the 10 nm displacement of the 555 nm band to the red observed for Hb⁵.

The distortion in the spectrum of α chain disappeared after a few microseconds. (When calculating the amplitude of spectral variation in this time range, correction was made for the partial (30-40%) recovery of the ligated proteins occurring during the first few hundred nanoseconds, due to geminate recombination^{5,8}.) A microsecond spectral variation was also observed for the β chain, although of very low amplitude. This variation showed that the Soret band of the photodeligated β chain is initially depressed by $\sim 1\%$ compared with that of the stable deoxy protein, a difference too small to appear in the transient spectrum shown in Fig. 1c.

The kinetics of the spectral evolution for the three proteins were obtained from measurements at 430 nm. Exponential decays showed that the transient form of the deligated proteins relaxes in a first-order reaction. The reactions occurring on excitation can then be written



The rate constant k is $4.1 \times 10^5\text{ s}^{-1}$ ($\pm 10\%$) for Hb, $6.6 \times 10^5\text{ s}^{-1}$ ($\pm 10\%$) for α and $6 \times 10^5\text{ s}^{-1}$ ($\pm 30\%$) for β . Activation energies (E_a) were determined from the temperature dependence of k (13-35 °C). Figure 2 shows the results in an Arrhenius plot. A value of $E_a = 9 \pm 1\text{ kcal mol}^{-1}$ was obtained for Hb and α . The results for the β chain, although only semi-quantitative, indicate that its E_a value is at least of the same order of magnitude as that of the other two proteins.

The spectral evolution (process (2)) is most probably due to changes in the haem environment occurring during structural relaxation of the proteins following photodeligation. Such a relaxation must occur in Hb because the reversible binding of ligand to this protein is known to be accompanied by a rearrangement of the four chains with respect to each other² (the quaternary change) and by tertiary conformation changes, including small readjustments of the haem-protein contacts¹¹⁻¹³. Structural studies of the isolated chains are lacking; however, X-ray investigations of myoglobin¹⁴ have shown that readjustments in the haem region occur on ligand binding to this non-cooperative protein and one may expect a similar behaviour for the isolated α and β chains.

As the non-cooperative isolated subunits show the evolution effect, the relaxation cannot be explained by the occurrence of a

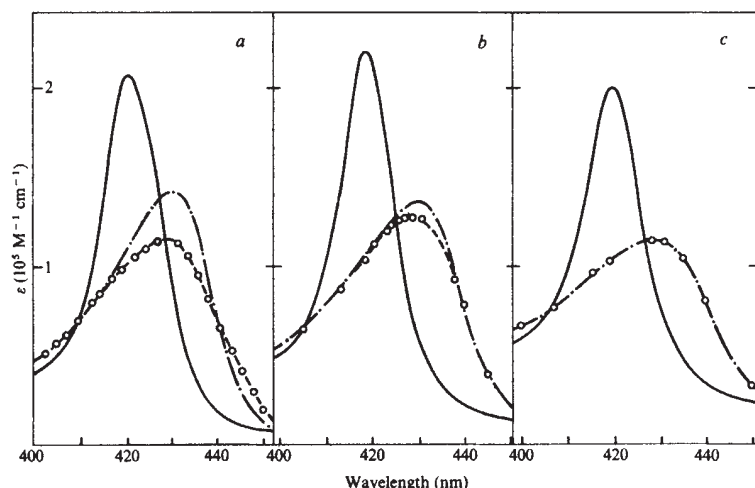


Fig. 1 Transient absorption spectra (○) of Hb (a), α (b) and β (c) measured after pulsed laser photodeligation of the corresponding CO complexes (end-of-pulse spectra). The absorption spectra of the CO-proteins (—) and the stable deoxy proteins (---) are also shown. Haem concentration $\sim 50 \mu\text{M}$.

quaternary structure change. However, it is striking that the spectral variation for Hb is not the mean of those for the isolated chains but is instead appreciably enhanced. This enhancement indicates that the α and β haem regions of this cooperative protein undergo more profound changes, probably because of the quaternary structure relaxation.

Gibson discovered in his early flash-photolytic studies¹⁵ the transient appearance of a spectrally distorted form of deoxy Hb with high reactivity ('quickly reacting form'). More recently, Sawicki and Gibson¹⁶ reported the transient spectrum observed following laser irradiation (of $1 \mu\text{s}$ duration) of HbCO. They did not detect any similar effect with the isolated chains and therefore stressed the role of the quaternary change in the transient Hb spectrum. Apart from the shape of this latter spectrum, our results differ from those of Sawicki and Gibson. However, differences in experimental conditions (time resolution, pH) preclude a more detailed comparison.

The E_a value of the spectral evolution observed in the present study is high, considering the weakness of the interactions determining the tertiary conformation. It is, on the other hand, of the same order of magnitude as the energy of the haem-proximal histidine (F8) coordination bond ($\sim 10 \text{ kcal mol}^{-1}$, as estimated from results for cobalttoporphyrins¹⁷). Also, the rate constant of process (2) is in the frequency range of the dissociation rate (10^5 – 10^6 s^{-1}) for a number of five-coordinate complexes of ferroporphyrins with nitrogenous bases^{18,19}. It is therefore tempting to suggest that the structural change in deoxy Hb (process (2)) is controlled by the thermally activated opening of the haem-histidine (F8) coordination bond. Removal of the stress in the unrelaxed deoxy Hb may be facilitated by haem and histidine movements during the transient opening of this bond.

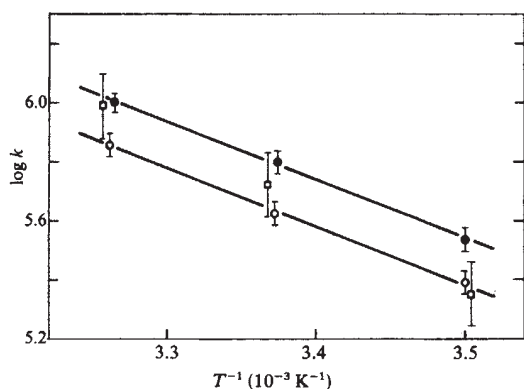


Fig. 2 Temperature variation of the rate constant k associated with the spectral evolution (process (2)) of photodeligated Hb (○), α (●) and β (□) on an Arrhenius plot. The high uncertainty in the points for the β subunit are due to the extremely small spectral change observed for this protein.

Received 14 July; accepted 4 November 1980.

1. Perutz, M. F., Muirhead, H., Cox, J. M. & Goaman, L. C. G. *Nature* **219**, 131–139 (1968).
2. Perutz, M. F. *Nature* **228**, 726–734 (1970).
3. Eisenberger, P., Shulman, R. G., Kincaid, B. M., Brown, G. S. & Ogawa, S. *Nature* **274**, 30–34 (1978).
4. Little, R. G. & Ibers, J. A. *J. Am. chem. Soc.* **96**, 4452–4463 (1974).
5. Alpert, B., El Mohsni, S., Lindqvist, L. & Tübel, F. *Chem. Phys. Lett.* **64**, 11–16 (1979).
6. Perutz, M. F. *J. Crystal Growth* **2**, 54–56 (1968).
7. Geraci, G., Parkhurst, L. J. & Gibson, Q. H. *J. biol. Chem.* **244**, 4664–4667 (1969).
8. Alpert, B., Lindqvist, L., El Mohsni, S. & Tübel, F. in *Interaction between Iron and Proteins in Oxygen and Electron Transport* (ed. Ho, C.) (Elsevier, Amsterdam, in the press).
9. Greene, B. I., Hochstrasser, R. M., Weisman, R. B. & Eaton, W. A. *Proc. natn. Acad. Sci. U.S.A.* **75**, 5255–5259 (1978).
10. Friedman, J. M. & Lyons, K. B. *Nature* **284**, 570–572 (1980).
11. Anderson, L. *J. molec. Biol.* **79**, 495–506 (1973).
12. Makinen, M. W. & Eaton, W. A. *Nature* **247**, 62–64 (1974).
13. Gelin, B. R. & Karplus, M. *Proc. natn. Acad. Sci. U.S.A.* **74**, 801–805 (1977).
14. Phillips, S. E. V. *Nature* **273**, 247–248 (1978).
15. Gibson, Q. H. *Biochem. J.* **71**, 293–303 (1959).
16. Sawicki, C. A. & Gibson, Q. H. *J. biol. chem.* **251**, 1533–1542 (1976).
17. Drago, R. S., Beugelsdijk, T., Breese, J. A. & Cannady, J. P. *J. Am. chem. Soc.* **100**, 5374–5382 (1978).
18. White, D. K., Cannon, J. B. & Traylor, T. G. *J. Am. chem. Soc.* **101**, 2443–2454 (1979).
19. Lavalette, D., Tétreau, C. & Momenteau, M. *J. Am. chem. Soc.* **101**, 5395–5401 (1979).

Human immunoglobulin variable region genes—DNA sequences of two V_κ genes and a pseudogene

David L. Bentley & T. H. Rabbitts

MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

The study of immunoglobulin genes at the molecular level can allow us to elucidate the origin of antibody diversity. Investigations of immunoglobulin gene structure in the mouse have shown that light chains are encoded by three gene segments: the C gene encoding the constant region and the V and J genes encoding the variable region. In antibody-producing cells the V and J genes join together to create a complete immunoglobulin gene^{1,2}. No data are available on the structure of human light chain variable region genes, but the variable regions of over 150 human κ light chain proteins have been sequenced and they comprise four distinct subgroups^{3–5}. Here we report the complete DNA sequences of three human κ variable region (V_κ) genes isolated from fetal liver DNA. The sequences demonstrate that two non-allelic genes encoding subgroup I proteins have more than 90% nucleotide homology in both coding and non-coding regions. Comparison of these human genes with two complete DNA sequences of mouse V_κ genes^{6,7} shows that V_κ gene structure is highly conserved between the two species, which suggests that V_κ genes rearrange during the differentiation of human lymphocytes by a very similar mechanism to that in the mouse^{8,9}. The sequence of a defective

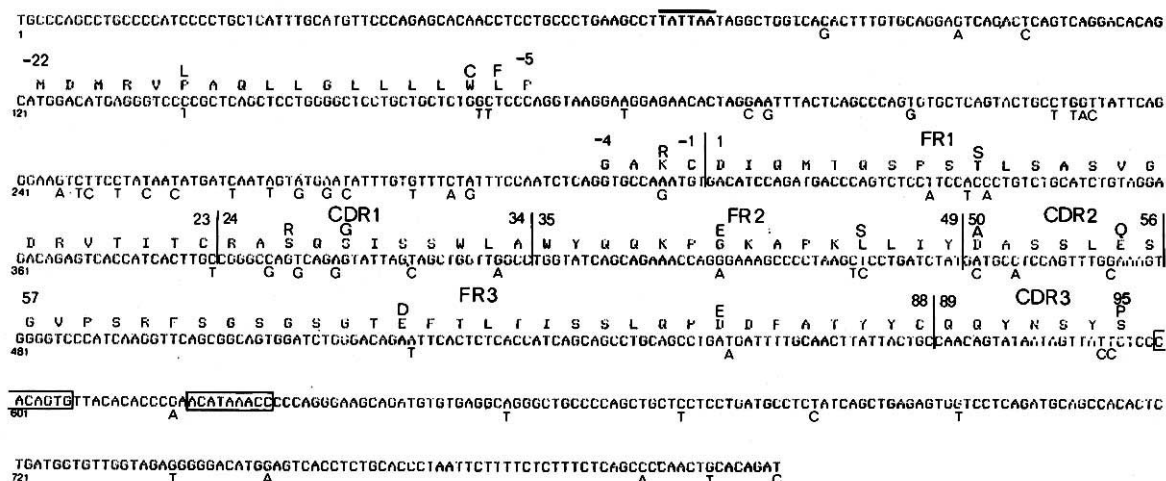


Fig. 2 Sequence of a second human V_{κ} gene, HK102. The HK102 gene was selected from the library of cloned human DNA fragments using as probe the HK101 *Pst*I fragment coding up to amino acid 79 subcloned in the plasmid pBR322. The HK102 V gene was sequenced by a rapid procedure in which recombinant λ phage DNA was digested with a restriction enzyme to yield a mixture of small (<500 base pairs) fragments which was then 'shotgun' cloned into an M13 vector treated with phosphatase to prevent self-ligation⁴². The resulting recombinants containing fragments of the variable region coding sequence were selected by plaque hybridization screening³⁶ using the HK101 V segment probe. *Sau*3A fragments were cloned in M13mp2/Bam⁴³ and *Alu*I fragments in M13mp73 (J. Messing, unpublished results). Use of two restriction enzymes made it possible to overlap all junctions between fragments except for the junction between *Alu*I fragments at position -15/-14 of the protein. These two fragments were aligned by analogy with the HK101 sequence. 58% of the sequence was obtained on both strands derived from independent M13 clones. The HK101 V_{κ} nucleotide and amino acid sequences are marked at points where they differ from HK102. A sequence related to the TATAAA box important for the initiation of transcription by RNA polymerase II in other genes⁴⁴ is overlined. The 'V-J joining sequences' are boxed.

and 3' sequences closely flanking mouse and human V_{κ} genes are as homologous as the coding regions. In the 130 bases 5' (upstream) of the first codon, HK101 and K2 are 72% homologous (excluding an insertion of 10 bases in the human sequence) and in the 90 bases 3' of the last codon they are 74% homologous. Further downstream no significant homology could be detected. The conservation of non-coding flanking sequences may suggest that they have functional significance for transcription or gene rearrangement but these predictions require experimental verification. It is apparent from the comparison of mouse and human V_{κ} genes that during evolution non-coding sequences diverge from one another in different ways from coding sequences. Divergent coding regions are related primarily by base substitutions, whereas non-coding regions also

contain insertions, deletions and duplications as observed for the mouse β -globin genes²⁸ and *Xenopus* vitellogenin genes²⁶.

As a first step towards a full description of diversity among germ-line V_{κ} genes we have isolated and sequenced two other human V_{κ} genes. The sequence of one of these, HK102, is shown in Fig. 2. HK101 and HK102 are probably non-allelic V_{κ} genes because we have identified restriction fragments corresponding to both of these in the DNA from each of 7 unrelated individuals using the Southern hybridization technique²⁹ (D.L.B. and T.H.R., in preparation). The HK101 and HK102 V_{κ} genes are more than 94% homologous throughout the coding region and the 5' and 3' flanking sequences (Fig. 2), and, as in the mouse-human comparison (Fig. 1), the intervening sequences are considerably less homologous (82%) than the

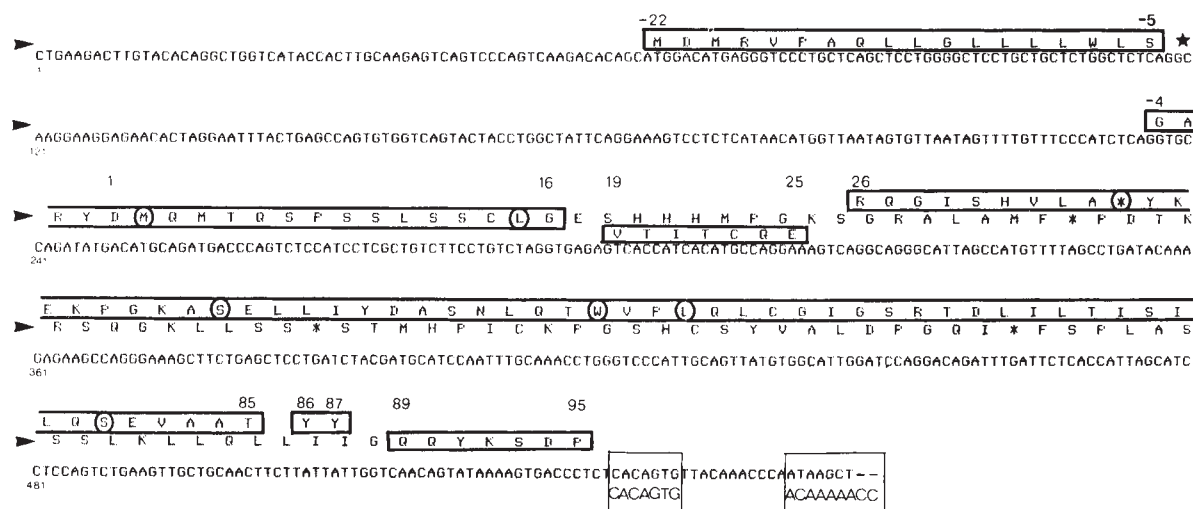


Fig. 3 The DNA sequence of a human V_{κ} pseudogene. Isolation of the gene, HK100, and the DNA sequencing strategy was as for HK102 (Fig. 2 legend). All junctions between *Alu*I and *Sau*3A fragments were sequenced. 92% of the sequence was obtained in both strands and the region where only one strand was sequenced was confirmed on two independent clones. The in-phase reading frame is indicated with arrows. The reading frames which correspond to normal immunoglobulin protein are in heavy boxes and amino acid positions are marked according to the standard numbering of normal κ light chains³. Asterisks mark termination codons and circles mark positions where highly conserved amino acids have been replaced due to apparent base-pair substitutions. The unconventional sequence at the 5' end of the intervening sequence is indicated by a star. Putative 'V-J joining sequences' are in light boxes. The final two blank positions refer to unidentified nucleotides beyond the *Alu*I fragment sequenced.

coding and flanking sequences. Both the HK101 and HK102 genes encode proteins that can be assigned to the V_{κ} 1b³⁰ subgroup but neither is identical to any known complete protein sequence. However, the HK102 sequence is closely related to the EU light chain protein³¹ from which it differs by 6 amino acids in the first 95. HK101 and HK102 differ at 14 amino acid residues in the signal peptide and in each of the framework and complementarity-determining regions of the protein. It seems likely that the extensive homology between these two human sequences is due to their recent common ancestry or to some gene correction mechanism³².

The sequence of a third human V_{κ} gene, HK100 (Fig. 3), is remarkable because it apparently cannot encode a functional immunoglobulin. The reading frame of this gene contains three in-phase termination codons which are a result of four small insertions and deletions in the DNA sequence relative to normal V_{κ} genes such as HK101 or HK102. These apparent mutations are a 2-base deletion at amino acid position 17/18, a 4-base insertion at 25/26, a 3-base insertion at 85/86 and a 1-base insertion in the invariant Cys codon at position 88. In addition, a number of apparent nucleotide substitutions in the coding region of HK100 change highly conserved amino acid residues at positions 2, 15, 44, 57, 60 and 80. A substitution at amino acid 35 replaces an invariant Trp codon, TGG, with a TGA nonsense codon (which is not in phase, however). Overall, the sequence of this abnormal gene and its flanking regions is 78% homologous to the HK101 sequence after maximizing homology⁴¹. The sequence of the HK100 V_{κ} gene suggests that it evolved by a mechanism typical of non-coding sequences, involving insertions and deletions as well as substitutions.

We do not know whether the HK100 V gene is capable of being transcribed but it is unlikely that a transcript could be correctly processed. The 5' end of the intervening sequence begins with a GC dinucleotide, not the highly conserved GT normally found at this position³³. This GC sequence has never been observed in 36 known functional RNA splicing sites³⁴.

Furthermore the HK100 V_{κ} gene may not be able to undergo V-J joining. Examination of the two sequence blocks downstream of the gene, thought to be important for V gene rearrangement, shows that the HK100 sequence possesses the first CACAGTG block but lacks a good match with the ACAAACCC consensus sequence for the second block. In fact it differs from the consensus at four of the first seven positions (see Fig. 1) whereas no other known V_{κ} gene differs from the consensus at more than one position³⁵.

Other genes apparently unable to encode functional products have been discovered for 5S rRNA and α and β globin (see ref. 10 for a review) and these have been termed pseudogenes. The HK100 V_{κ} gene is the first example of an immunoglobulin variable region pseudogene.

We thank Dr T. Maniatis for providing the library of cloned human DNA and Drs Perry, Marcu and Schibler for the plasmid p κ (11)24 (ref. 16), J. Karn for Sau3A, and B. J. Meyer, J. Sims and S. Fields for critical reading of the manuscript.

Received 16 October; accepted 24 November 1980.

- Brack, C. *et al.* *Cell* **15**, 1-14 (1978).
- Seidman, J. G. & Leder, P. *Nature* **276**, 790-795 (1978).
- Kabat, E. A. *et al.* *NIH Publ. No.* 80-2008 (1979).
- Milstein, C. *Nature* **216**, 330-332 (1967).
- Wang, A. C. *et al.* *Nature new Biol.* **243**, 126-127 (1973).
- Seidman, J. G. *et al.* *Nature* **280**, 370-375 (1979).
- Nishioka, Y. & Leder, P. *J. biol. Chem.* **255**, 3691-3694 (1980).
- Sakano, H. *et al.* *Nature* **280**, 288-294 (1979).
- Max, E. E. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **76**, 3450-3454 (1979).
- Proudfoot, N. J. *Nature* **286**, 840-841 (1980).
- Jahn, C. L. *et al.* *Cell* **21**, 159-168 (1980).
- Lacy, E. & Maniatis, T. *Cell* **21**, 545-553 (1980).
- Proudfoot, N. J. & Maniatis, T. *Cell* **21**, 537-544 (1980).
- Jacq, C., Miller, J. R. & Brownlee, G. G. *Cell* **12**, 109-120 (1977).
- Rabbitts, T. H., Matthysens, G. & Hamlyn, P. H. *Nature* **284**, 238-243 (1980).
- Schibler, U., Marcu, K. B. & Perry, R. P. *Cell* **15**, 1495-1509 (1978).
- Lawn, R. M. *et al.* *Cell* **15**, 1157-1174 (1978).
- Blattner, F. *et al.* *Science* **196**, 161-169 (1977).
- Gronenborn, B. & Messing, J. *Nature* **272**, 375-377 (1978).
- Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463-5467 (1977).
- Schreier, P. H. & Cortese, R. *J. molec. Biol.* **129**, 169-172 (1979).
- Matthysens, G. & Rabbitts, T. H. *Proc. natn. Acad. Sci. U.S.A.* (in the press).
- Early, P. *et al.* *Cell* **19**, 981-999 (1980).

- Bernard, O. *et al.* *Cell* **15**, 1133-1144 (1978).
- van den Berg, J. *et al.* *Nature* **276**, 37-44 (1978).
- Wahli, W. *et al.* *Cell* **20**, 107-117 (1980).
- Perier, F. *et al.* *Cell* **20**, 555-566 (1980).
- Konkel, E. *et al.* *Cell* **18**, 865-873 (1979).
- Southern, E. M. *J. molec. Biol.* **98**, 503-517 (1975).
- Milstein, C. P. & Deverson, E. V. *Biochem. J.* **123**, 945-958 (1971).
- Gottlieb, P. D. *et al.* *Biochemistry* **9**, 3155-3161 (1970).
- Tartof, K. D. *An. Rev. Genet.* **9**, 355-385 (1975).
- Breathnach, R. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **75**, 4853-4857 (1978).
- Lerner, M. R. *et al.* *Nature* **283**, 220-224 (1980).
- Sakano, H. *et al.* *Nature* **286**, 676-683 (1980).
- Benton, W. D. & Davis, R. W. *Science* **196**, 180-182 (1977).
- Maniatis, T., Jeffrey, A. & Kleid, D. G. *Proc. natn. Acad. Sci. U.S.A.* **72**, 1184-1188 (1975).
- Smith, H. O. & Birnstiel, M. L. *Nucleic Acids Res.* **3**, 2387-2398 (1976).
- Backman, K., Ptashne, M. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **73**, 4174-4178 (1976).
- Anderson, S. *et al.* *Nucleic Acids Res.* **8**, 1731-1745 (1980).
- Staden, R. *Nucleic Acids Res.* **6**, 2601-2610 (1979).
- Ullrich, A. *et al.* *Science* **196**, 1313-1319 (1977).
- Rothstein, R. J. *et al.* *Meth. Enzym.* **68**, 98-109 (1979).
- Corden, J. *et al.* *Science* **209**, 1406-1414 (1980).

Dimethyl sulphoxide reduces anti-receptor antibody titres in experimental myasthenia gravis

Alan Pestronk & Daniel B. Drachman

Johns Hopkins University, School of Medicine, Department of Neurology, 1721 E. Madison Street, Baltimore, Maryland 21205

The basic abnormality in myasthenia gravis (MG) is a reduction of acetylcholine receptors (AChRs) at neuromuscular junctions, due to an autoimmune attack directed against the receptors¹. Several lines of evidence support the idea that humoral immune mechanisms play an important part in this process: anti-AChR antibodies are present in the sera of more than 80% of myasthenic patients²⁻⁴; the pathogenicity of the antibodies has been demonstrated by passive transfer of IgG from myasthenic patients to mice which reproduces the typical features of MG in the recipient animals^{5,6}; finally, procedures that reduce levels of anti-AChR antibody, such as plasmapheresis, produce improvement in the clinical symptoms of MG⁷. Present treatments of MG depend largely on the prolonged use of immunosuppressive drugs, which have significant drawbacks^{1,8-10}, for they usually produce only a slow fall in antibody titres, and have toxic side effects. An agent that could produce a rapid and sustained fall of autoantibody titre might have wide application in the treatment of MG and other autoimmune diseases. In the course of testing immunosuppressive drugs for their effect in treating experimental autoimmune MG (EAMG), we discovered that the vehicle used to dissolve some of these agents, dimethyl sulphoxide (DMSO), itself produced a rapid and sustained fall of anti-AChR antibody titre. We now report similar results for a controlled trial of DMSO treatment in rats with EAMG.

EAMG was produced in female Lewis rats (200-225 g) by immunizing them with 50 μ g of AChR, emulsified in an equal volume of Freund's complete adjuvant, injected intradermally. AChR was extracted from the electric organs of *Torpedo californica* with Triton X-100 and purified by affinity chromatography, using α -toxin from the venom of the cobra *Naja naja siamensis* covalently linked to Sepharose 4B beads as the binding ligand¹¹.

Measurement of anti-AChR antibodies was carried out by a double precipitation radioimmunoassay like those previously described^{2,4}. Antibodies directed against *Torpedo* AChR were measured using as the antigen 0.5×10^{-12} moles of purified *Torpedo* AChR coupled to ¹²⁵I- α -bungarotoxin (¹²⁵I- α -BTX). Serial dilutions of sera from the experimental rats were used, and precipitation was produced by the addition of goat anti-rat IgG. The antibody titre, determined from the linear portion of the dilution curve, is expressed as moles $\times 10^{-9}$ of α -BTX precipitated per litre. For determinations of antibody against rat AChR, an aliquot of Triton-extracted denervated rat muscle containing 0.5×10^{-12} moles of AChR was used as the antigen,

coupled to ^{125}I - α -BTX as above. These assays were sensitive to antibody concentrations of 0.5×10^{-9} moles per l of serum. Determinations of antibodies against both *Torpedo* AChR (the immunizing antigen) and rat AChR (the 'self' antigen) were made repeatedly up to 2 months after the start of the experiment. Levels of rat IgG were measured in sera from treated and control EAMG animals by a commercial radial immunodiffusion assay (Hyland).

The original purpose of our investigations was to examine whether frentizole, a drug thought to be particularly effective in suppressing humoral immune responses, would reduce the levels of anti-AChR antibody in rats with EAMG. Because frentizole is poorly soluble in water it was administered in DMSO solution; control EAMG animals received DMSO alone. Our results showed that anti-AChR titres fell by 50% within 1 week in both experimental and control animals.

A therapeutic trial of DMSO alone was therefore undertaken to determine whether DMSO was responsible for the fall in the antibody titre. Experimental rats with EAMG were treated with a 2-week course of daily intraperitoneal injections of DMSO (1.0 ml) starting 1 month after immunization with AChR. Control EAMG animals immunized at the same time were given injections of physiological saline.

After 1 week of treatment with DMSO the titre of anti-*Torpedo* AChR antibody decreased to a mean of 46% of pretreatment levels. The titres remained markedly reduced, reaching 32% of pretreatment levels 6 weeks after discontinuation (Fig. 1a). In contrast, antibody titres in control animals remained elevated, eventually rising to a level 30% higher than that at the start of the study (Fig. 1a). There was also a rapid reduction in levels of the autoantibody, against rat AChR. One week after the start of DMSO treatment anti-rat AChR antibody titres had fallen to 48% of pretreatment levels (Fig. 1b). By the end of the study, the mean anti-rat AChR antibody titre was 14% of the pretreatment level, and 12% of the level in untreated controls. In control animals autoantibody titres remained relatively constant, averaging 5% higher at the end of the study. Serum concentrations of IgG were initially reduced by DMSO treatment, being 42% of pretreatment levels after 1 week. However, in contrast to the anti-AChR antibody, IgG rapidly returned to pretreatment levels (Fig. 2). By 2 weeks after the end of the treatment course, the mean IgG level was not significantly different from the levels in pretreatment or control animals.

We carried out additional control experiments which showed that DMSO did not interfere with the assay for anti-AChR antibody. Antibody titres were not significantly affected by levels of DMSO of up to 0.5% in serum. Further, antibody titres

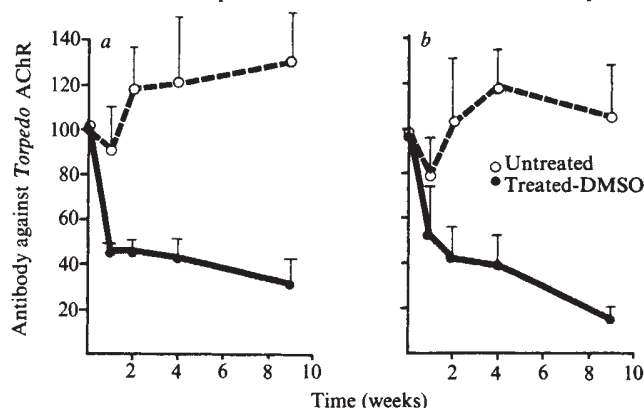


Fig. 1 a, Antibody titres against *Torpedo* AChR; b, antibody titres against rat AChR. Dashed line indicates titres in untreated EAMG rats, solid line those in DMSO-treated EAMG rats. Treatment was given at time 0 (4 weeks after induction of EAMG). Antibody titres at time 0 were normalized to 100%. The absolute values of antibody at this time were $640\text{--}2,300 \text{ mol} \times 10^{-9} \text{ l}^{-1}$ against *Torpedo* AChR and $5\text{--}60 \text{ mol} \times 10^{-9} \text{ l}^{-1}$ against rat AChR. Subsequent values represent the mean of percentages of the initial titre (from eight treated and seven untreated rats) $\pm$ s.e.

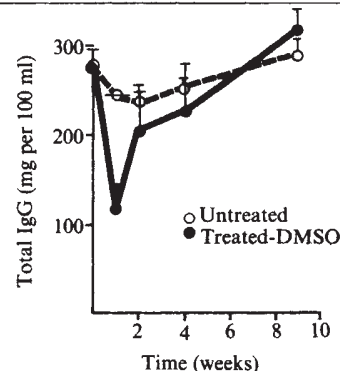


Fig. 2 Levels of total IgG. Dashed line indicates levels in untreated EAMG rats, solid line those in DMSO-treated EAMG rats. Treatment was given at time 0 (4 weeks after induction of EAMG).

in EAMG animals treated 8 h previously with a single dose of DMSO did not differ from titres in untreated animals.

Our results show that treatment with DMSO reduced antibody titres against both foreign (*Torpedo*) and self (rat) AChR. The fall in titres was rapid and sustained, and titres did not rise even after DMSO was discontinued. The mechanism by which DMSO reduces antibody titres is unknown. The fall in anti-AChR antibody titres is not attributable to a direct effect of DMSO on the antibody, as there was no change in the titre after acute treatment with DMSO. The effect is not due to interference with the antibody assay, because relatively high serum concentrations of DMSO had no effect on antibody measurement. In addition, the half life of DMSO is less than 1 day¹² and anti-AChR antibody titres remained low for at least 6 weeks after the cessation of treatment. A more likely explanation for the fall in titres is that DMSO reduced the production of antibody by lymphoid cells. It remains to be determined whether this effect is due to a direct action of DMSO on antibody-producing cells or to an indirect effect on immune regulatory mechanisms such as suppressor cells. In any case, active immune responses may be especially susceptible to DMSO, for anti-AChR antibody titres remained low, whereas the total IgG concentration returned to normal.

It is encouraging that the fall of anti-AChR antibody titres after DMSO treatment is comparable to that which results from plasmapheresis. This suggests that DMSO might be effective in the treatment of diseases such as MG where antibodies have an important pathogenic role¹ and plasmapheresis provides only temporary improvement⁷. The finding that at the longer times after treatment, autoantibodies against AChR remained depressed whereas the total levels of IgG returned to normal may be relevant to the potential usefulness of DMSO in human disease states.

Thus, we have found that DMSO treatment produces a rapid and sustained fall in anti-AChR antibody in EAMG. It seems likely that DMSO might also be effective in ameliorating other active humoral immune responses. If so, DMSO may provide an effective new treatment for immune diseases mediated by autoantibodies.

This work was supported by grants from the NIH and the Muscular Dystrophy Association of America.

Received 7 August; accepted 24 October 1980.

- Drachman, D. B. *New Engl. J. Med.* **298**, 136–142, 186–193 (1978).
- Lindstrom, J. M., Seybold, M. E., Lennon, V. A. & Whittingham, S. *Neurology* **26**, 1054–1059 (1979).
- Mittag, T., Kornfeld, P., Tormay, A. & Woo, C. *New Engl. J. Med.* **294**, 691–694 (1976).
- Appel, S. H., Almon, R. R. & Levy, N. *New Engl. J. Med.* **293**, 760–761 (1975).
- Toyka, K. V. *et al.* *New Engl. J. Med.* **296**, 125–131 (1977).
- Howard, J. F. & Sanders, D. B. *Neurology* **30**, 760–764 (1980).
- Dau, P. C. (ed.) *Plasmapheresis and Immunobiology of Myasthenia Gravis* (Houghton Mifflin, Boston, 1979).
- Hertel, G., Mertens, H. G., Ruether, P. & Ricker, K. in *Plasmapheresis and the Immunobiology of Myasthenia Gravis* (ed. Dau, P. C.) Ch. 29 (Houghton Mifflin, Boston, 1979).
- Sanders, D. B., Howard, J. F. Jr, Johns, T. R. & Campa, J. F. in *Plasmapheresis and the Immunobiology of Myasthenia Gravis* (ed. Dau, P. C.) 289–306 (Houghton Mifflin, Boston, 1979).
- Rowland, L. P. *Trans. Am. neurol. Ass.* **133**, 277–302 (1978).
- Eldefrawi, M. E. & Eldefrawi, A. T. *Archs Biochem. Biophys.* **159**, 362–373 (1973).
- Hucker, H. B., Ahmad, P. M. & Miller, E. A. *J. Pharmac. exp. Ther.* **154**, 176–184 (1966).

Anomalous region on Mars: implications for near-surface liquid water*

S. H. Zisk

NEROC Haystack Observatory, Westford, Massachusetts 01886

P. J. Mouginis-Mark

Department of Geological Sciences, Brown University, Providence, Rhode Island 02912

An anomalous region has been identified on Mars from the 1971 and 1973 Earth-based Goldstone radar data. This region is characterized by coincident very high radar reflectivities, and unusual smoothness. Very few realistic surface morphologies can generate such return-signals. Liquid water within ~50–100 cm of the surface is one possibility, an interpretation strengthened by an apparent seasonal variation in surface reflectivity for the same locality.

ANALYSIS of radar measurements of the martian surface enables textural-compositional differences of materials to be interpreted on a scale unobtainable from orbital images^{1–3}. Estimates of the dielectric constant (from inherent reflectivity) and small-scale slope statistics (see, for example, the *C*-factor analysis developed by Hagfors⁴) can be used to characterize surface units⁵ or to delineate areas with unusual material properties.

An extensive data base was compiled from Earth-based radar measurements of the martian surface by Downs *et al.*² in 1971 and 1973 at the Jet Propulsion Laboratory's Goldstone Facility. This combined set of observations gave almost complete coverage between 14 and 22° S. The high dielectric constant of liquid water compared with both ice and dry geological materials⁶ means that radar observations should provide excellent evidence for or against the existence of water on Mars. In particular, this radar data set can provide additional evidence regarding the existence of near-equatorial regions of H₂O outgassing ('oases') proposed by Huguenin *et al.*^{7,8}. Liquid water should affect the radar characteristics of the surface, if any overburden is relatively transparent to the radar energy. If the water were allowed to form a free surface, as opposed to an adsorbed layer on a mineral substrate, then the radar-reflecting facets would be expected to be nearly horizontal due to gravitational forces. Hence this type of surface would give a radar return with a very high smoothness as well as high reflectivity. We report here our attempts to search for anomalous areas within the radar data set, where both high reflectivity and *C*-factor values could be attributed to a layer of liquid water or damp soil.

Experimental method

The radar determination of surface smoothness and reflectivity is derived from the observed variation of echo strength as a function of local incidence angle^{4,6}. The radar echo at any incidence angle is the incoherent sum of reflections from surface facets within the resolution cell which are oriented perpendicular to the radar beam. Strictly speaking, the reflecting surfaces are the Fresnel zones of these elements, so that the cross-section of each surface element depends on its radius of curvature or, where the radius is large, on its area⁶. The size of the radar

resolution cell is much larger than even the planetary-radius Fresnel zone. It is fixed by the resolution of the radar signal in time-delay and Doppler frequency, which respectively determine the resolution along the projected axis of rotation and at right angles to it. For the radar data considered here, the resolution cell measures ~80 km by 10 km, compared with the 0.9-km size of the Fresnel zone for the planet's radius of curvature².

The *C*-factor parameter used to describe the surface smoothness quantitatively was derived by Hagfors⁴. He postulated a set of models for surface topography (based on observations of the Moon), for each of which he calculated the number distribution of surface-facet slopes, and thence the expected curve of radar strength compared with incidence angle (*C*-factor template). The radar-observed curve was computer-fitted to these hypothetical curves to determine the best-fit value for the *C*-factor. For areas with moderate to high smoothness (*C*-factor >> 100), the expression for the r.m.s. surface slope a_{rms} is

$$a_{rms} = C^{-1/2} \text{ rad}$$

For real surfaces which do not match these models, the measured curve of echo strength compared with the incidence angle will not agree well with any of the templates, and a large standard error will result from the fitting process. Smooth, random surfaces will generally show a good fit to the templates, because there is little dependence on the details of the model. Surfaces with high relief, on the other hand, cannot usually be described well by only one parameter. Also, a non-random surface such as an area of sand-dunes or a cluster of comparable-size craters, cannot be described well by the standard *C*-factor parameter at any observable amplitude of relief.

The morphology of the surface determines the shape of the curve and hence the value of *C*-factor. The amplitude of the curve is a function only of the reflectivity of the material of which the surface is composed, and is another independent output from the template-fitting process.

In the present radar measurements, the same surface-resolution element was observed at various incidence angles (by using the Doppler resolution in longitude) to form the curve of echo strength versus incidence angle. There is therefore no implicit

* **Erratum:** Due to an error in the *Nature* office, the original rather than the subsequently revised version of this article appeared in the issue of 13 November (288, 126–129, 1980). We here reprint the article in its correct form.

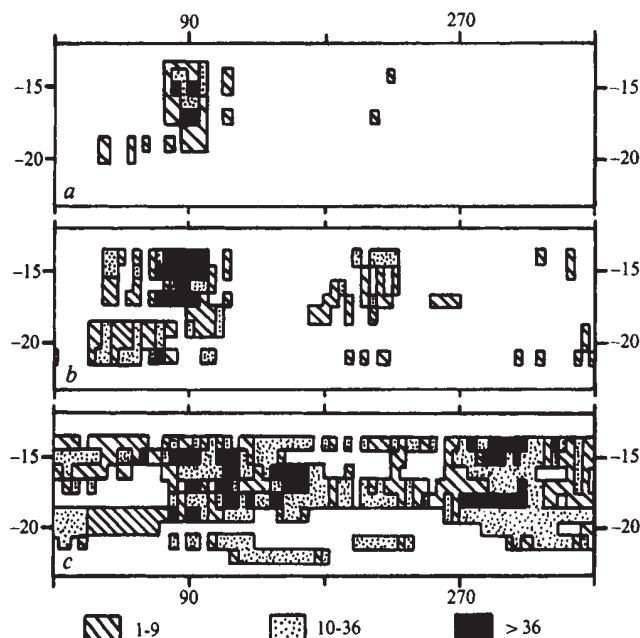


Fig. 1 Number density of radar sample points with given characteristics per 1° latitude, 5° longitude sample bins. *a*, Reflectivity $>9\%$ and *C*-factor $>3,200$ coincidentally; *b*, *C*-factor $>3,200$; *c*, reflectivity $>9\%$. Note that although data points with high reflectivity or *C*-factor occur over a wide area of Mars, coincident high values in both parameters are found exclusively at the 'oasis' areas proposed by Huguenin *et al.*^{7,8} (Solis Lacus, at $15\text{--}30^\circ$ S, $80\text{--}90^\circ$ W; and Noachis-Hellespontus, at $15\text{--}25^\circ$ S, $305\text{--}315^\circ$ W).

assumption of large-area uniformity of the surface, as was the case for some earlier measurements. Any systematic variation in surface composition as, for example, a low-reflectivity mantled surface interspersed with high-relief, high-reflectivity rock formations, will show up as a poor fit (high standard error) to the *C*-factor templates⁵. A mixture of high- and low-reflectivity materials exhibiting the same morphology will yield a low-error *C*-factor value, with an intermediate reflectivity. Other surface configurations can be hypothesized and their radar-measured parameters calculated. The reverse calculation, however, from the radar data to the surface configuration, cannot be carried out unambiguously for all combinations of reflectivity and *C*-factor. As the surface becomes less smooth, the variety of possible configurations increases and the reliability of the template-fitting becomes much worse, partly because of the reduced strength of the widely-scattered signal and partly because of the above-mentioned variety of possible surface topographies.

Analysis

The data were scanned for regions with a coincident high reflectivity and smoothness. Limits of 9% reflectivity and $3,200$ *C*-factor (r.m.s. slopes $<1^\circ$) were chosen arbitrarily. Variations from 9 to 11% and from $2,400$ to $3,200$ respectively had only a slight effect on the shape of the anomalous area and on the number of observation points. For these measurement values, the formal errors (standard deviations) were $\sim 10\%$ of the data value.

Figure 1*a* illustrates the areocentric position and number densities of these high-reflectivity, high-smoothness measurements. Only two regions on the planet share this combined radar characteristic. The most prominent is a region centred at 16° S 90° W. There is also a pair of smaller areas to the north-west of Hellas. These two regions are remarkably close to the Solis Lacus (25° S, 85° W) and Noachis-Hellespontus (20° S, 310° W) oases proposed by Huguenin *et al.*^{7,8}.

For comparison, Fig. 1*b,c* indicates the areas having only their reflectivity or *C*-factor above the defined limits. Note that there are many areas of high reflectivity, indicating a profusion of well-compacted or rocky soils on the martian surface. Areas of high smoothness are less common, although they are not restricted to the anomalies described here.

As an argument for the reality of the anomalous areas, note that the map of Fig. 1*a* includes radar measurements from both the 1971 and 1973 conjunctions. Almost no anomalous data points occur elsewhere on the planet, despite the near-global coverage of the data set between 14 and 22° S. In the 300×300 km anomaly itself, on the other hand, there are several hundred points.

Seasonal variability

Additional characteristics of Solis Lacus can be determined by comparing radar data derived for nearly coincident ground tracks at different times of the martian year. Note that practical limitations of Earth-based radar systems dictate that radar observations of the sub-Earth region must be made near opposition, and hence near local martian summer. Winter measurements would at best be extremely difficult to make using Earth-based radar.

A combination of data from the 1971 and 1973 oppositions provides an example of overlapping radar swaths: latitudes 15.45° S, 15.43° S, and 15.81° S were sampled between 80 and 100° W during the seasonal range from early to middle local martian summer ($L_s = 258^\circ$, 273° and 294° respectively). Note that, as L_s represents the longitude of the Sun as seen from Mars, the difference between $L_s = 258^\circ$ and 294° is roughly equivalent to the shift from early June to mid-July on Earth (Northern Hemisphere). Because of the nearly identical locations of these radar profiles (the north-south extent of the radar resolution cell is 1.3° , giving an overlap of at least 70% between swaths), the reflectivity curves for identical surfaces should be very similar. However, Fig. 2 indicates that there is a discrepancy in the reflectivity between $L_s = 258^\circ$ and 273° . The 273° profile indicates an increase of $1.0\text{--}1.5\%$ in the reflectivity compared to the same longitude at $L_s = 258^\circ$, while the 294° profile shows a further enhancement of $2.5\text{--}3.5\%$. For large parts of the region between 82 and 100° W, there is a systematically higher reflectivity with advancing L_s , that is season. Such a

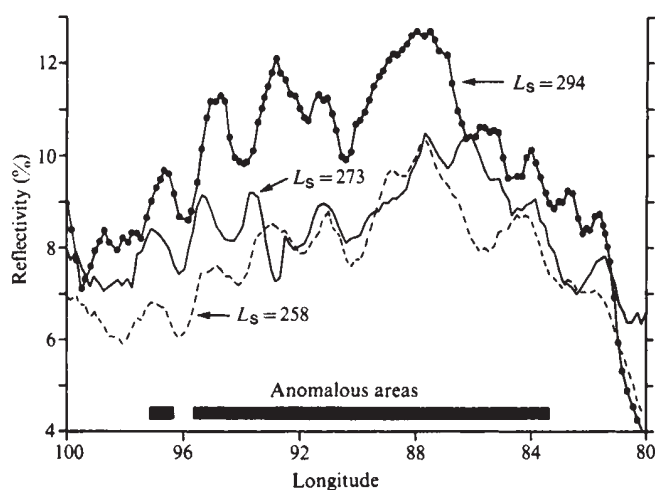


Fig. 2 Temporal variations in radar reflectivity for an area of Solis Lacus at $15.43\text{--}15.81^\circ$ S. Curves represent a running 5-point average at 0.16° longitude intervals. Radar data was taken during 1971 ($L_s = 258^\circ$) and 1973 ($L_s = 273^\circ$ and 294°) Mars oppositions at latitudes 15.44 , 15.43 and 15.81° S respectively. The 'anomalous areas' refer to parts of the ground track at $L_s = 294^\circ$ where reflectivity is $>9\%$ and *C*-factor $>3,200$ (see also Fig. 4*a*).

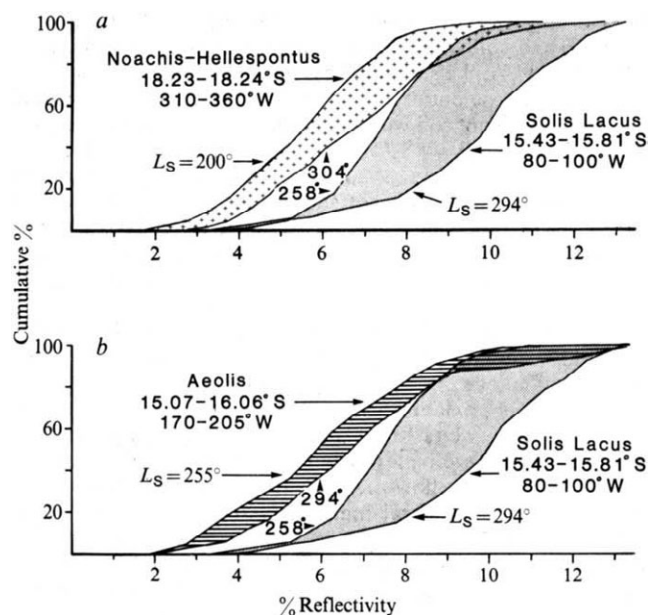


Fig. 3 The seasonal variations in radar reflectivity compared here for three regions of Mars. The horizontal axis is reflectivity and the vertical axis is the percentage of data points having less than the indicated reflectivity. *a*, Compares the strongly anomalous Solis Lacus area with the weakly anomalous Noachis-Hellespontus region. *b*, Compares Solis Lacus with an area in Aeolis, which is taken to be representative of Mars as a whole and shows no anomaly. Note that all three areas display an increase in percentage reflectivity with increasing martian season, but that Solis Lacus has a much more pronounced seasonal change, in addition to its higher absolute radar reflectivity.

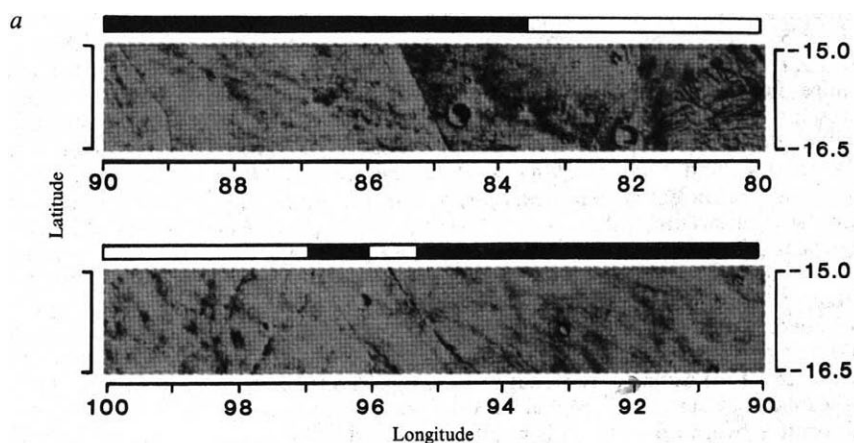
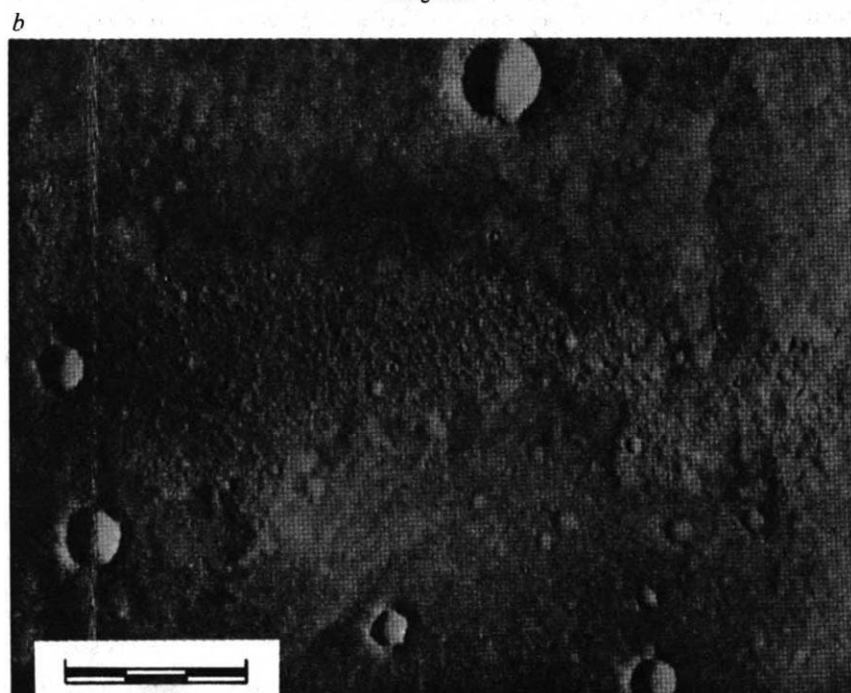


Fig. 4 Viking Orbiter Images of the anomalous radar areas. *a*, Ground track shown in Fig. 2; *b*, high resolution for an area in Solis Lacus at 18.7° S. Note that the images do not support a hypothetical smooth, mantle-free surface as an explanation of the radar anomaly. Numerous 10-km craters and lava flow fronts are in evidence (USGS photomosaics I-1183 and I-1190; and Viking frame 773A10). Scale bar, 3 km.



phenomenon seems to be real, since some parts of the data (for example, between 83–84° W and 88–92° W for $L_s = 258^\circ$ and 273°) show no difference in reflectivity despite their respective acquisition dates in 1971 and 1973. This is strong evidence for the adequacy of the calibration of the original radar data, and also for the lack of any temporal variation in the data from instrumental causes.

In another test for possible systematic effects, we are searching the radar data at areas other than the observed anomaly for quasi-seasonal variability. Some of the results are depicted in Fig. 3. Figure 3a presents a plot of the range of variability of the reflectivity with L_s for Noachis-Hellespontus, which contains the only cluster of anomalous points outside the Solis Lacus region. Figure 3b presents Aeolis as an example of a non-anomalous region. Figure 3 shows that there is a 'seasonal' variability in much of the observed area, in the same sense (increased reflectivity with advancing summer); but that Solis Lacus exhibits the greatest range by more than a factor of 2 of those areas so far examined. The general variability may indicate scattered expressions of near-surface water and/or permafrost; or possibly a low-level, systematic and time-dependent flaw in the calibration.

Discussion

Measurements showing a high radar reflectivity (high dielectric constant) would be expected not only for water surfaces, but also for large areas of smooth rock within 1 m of the surface. In these conditions, anorthositic rock, with a dielectric constant K of about 5, would have a reflectivity of about 15% (basalt, with $K = 9$, is >25%; water, with $K = 81$, is >65%) if it covered the entire surface area of the radar cell (~750 km²). Figure 4 presents Viking Orbiter images of the same area as Fig. 2. Although few craters larger than ~10 km diameter are within the radar ground track, many moderately well preserved lava flow fronts can be easily identified. Morphologically comparable lava flows elsewhere on Mars either have radar reflectivities of 2–6% or C -factors <2,000 (refs 5, 9, 10). Unless the small-scale surface texture of the Solis Lacus region is unrelated to its large-scale appearance in Fig. 3, the hypothesis of an extensive rock pediment is untenable as an explanation for the high values of radar reflectivity and C -factor.

Moreover, to explain a seasonal variability in the reflectivity of a solid rock surface, one must hypothesize, for example, an absorptive overburden which is regularly scoured away as the summer progresses. The thickness of overburden required for the observed changes in dielectric constant have been calculated on the basis of published dielectric constants and absorption lengths for basaltic rock and loosely-compacted rock powder¹¹. Although resonant effects of thin highly-planar soil and rock layers might be observed in laboratory conditions, such conditions do not exist on a planetary surface. Consequently, the reflectivity was calculated for a random-thickness soil layer masking a solid rock surface. Assuming dielectric constants of 2.0 for soil and 8.0 for rock, the net reflection coefficient for bare rock is 23%; for a soil layer at least 14 cm (one wavelength) thick, is 11%; and drops to 6% for a soil depth of 1.4 m. For lunar rock chemistry rather than terrestrial rock chemistry, the 6% reflectivity would require more than 10 m of rock-free soil. The problem is complicated by other factors such as variation in

soil porosity with depth, but the conclusion, nevertheless, is that for rock and soil surface to exhibit a measured reflectivity of more than 11%, the soil thickness must be in the range from 10 to 20 cm over the entire area; and that an additional deposit of the order of 1 m of soil would be required to reduce further the dielectric constant to 8%. Although there is undoubtedly a seasonal redistribution of martian dust, photographs at the Viking Lander sites¹² do not show the buildup of the tens of centimetres of dust that would be needed to effect a change from 11 to 8% in the average radar reflectivity. Alternative explanations, such as seasonal changes in the bulk chemistry of the near-surface rocks, seem also to be unlikely.

Conversely, a transition from frozen water (reflectivity >7%) to liquid water (reflectivity >64%) would produce just such a striking difference in the observed radar reflectivity. Considerable supporting evidence exists that permits the presence of liquid water on Mars to be hypothesized. The large martian channel systems are believed to be the products of large-scale release and surface flow of water from confined aquifers^{13,14}. Thermal and IR-spectral measurements of the North Polar cap¹⁵ indicate abundant water ice, while 1–3 wt% of adsorbed water was detected in samples taken by the Viking Landers¹⁶. In addition, on a smaller scale, both spectrophotometer measurements made during the 1973 dust storm¹⁷ and measurements from the Viking water-vapour mapping experiment¹⁸ indicate that a considerable amount of H₂O outgassing occurs at present within Solis Lacus. Although the surface temperatures at the Viking Lander sites are only in the range 230–250 K (ref. 19), for an assumed analogy with terrestrial geology the water reservoir could be highly saline due to the salts leached from the surrounding rocks²⁰. For Mars, saline solutions have been shown to be stable in their liquid phase in current surface conditions at temperatures as low as 210 K (ref. 21).

We conclude, therefore, that only surface materials containing liquid water are consistent with: (1) the coincident very high radar reflectivities and C -factor values; (2) the photographically observed surface morphology; and (3) the increase in reflectivity values from spring to summer. Some inferences can also be made about the regional extent of the near-surface expression of the liquid water at Solis Lacus. An extensive liquid water surface (such as a lake) would have a much higher reflectivity than the observed maximum value of about 16%. More likely is a reduction in the reflectivity by an overlying layer of soil, as outlined above for the rock-soil combination, or, alternatively (or additionally), the liquid water might be observable only through random patches of the surface because of variations in the thickness or density of the overburden. In the latter case, the observed reflection values would be from a mixture of high and low reflectivity surface elements. In this case, the C -factor measurements would depend mainly on the highly reflecting regions (the oases) rather than the dry intervening areas. In either case, we are convinced by the radar data that liquid water does indeed exist near the surface of Solis Lacus during Southern Hemisphere summer.

We thank George Downs for radar data, and Richard Simpson and Gordon Pettengill for helpful comments. This research was partly supported by NASA under grants NSG-7179, NSG-7569, and NSG-7570. Haystack Observatory is operated by MIT under NSF grant AST-7920168.

Received 7 April; accepted 13 October 1980.

- Pettengill, G. H., Shapiro, I. I. & Rogers, A. E. E. *Icarus* **18**, 22–28 (1973).
- Downs, G. S., Reichley, P. E. & Green, R. R. *Icarus* **26**, 273–312 (1975).
- Simpson, R. A., Tyler, G. L. & Lipa, B. J. *Icarus* **32**, 147–167 (1977).
- Hagfors, T. *J. geophys. Res.* **69**, 3779–3784 (1964).
- Mouginis-Mark, P. J. & Zisk, S. H. *Lunar planet. Sci.* **11**, 771–773 (1980).
- Evans, J. V. & Hagfors, T. *J. Radar Astronomy* (McGraw-Hill, New York, 1968).
- Huguenin, R. L., Clifford, S. M., Sullivan, C. A. & Miller, K. J. *NASA Tech. Mem.* **80339**, 208–214 (1979).
- Huguenin, R. L. & Clifford, S. M. *Bull. Am. astr. Soc.* **11**, 580 (1979).
- Mouginis-Mark, P. J. & Zisk, S. H. *Lunar planet. Sci.* **11**, 768–770 (1980).
- Mouginis-Mark, P. J., Cintala, M. J. & Whitford-Stark, J. L. *Lunar planet. Sci.* **11**, 762–764 (1980).

- Campbell, M. J. & Ulrichs, J. *J. geophys. Res.* **74**, 5867–5880 (1969).
- Arvidson, R., Guinness, E. & Lee, S. *Nature* **278**, 533–535 (1979).
- Baker, V. R. *J. geophys. Res.* **84**, 7985–7993 (1979).
- Carr, M. H. *J. geophys. Res.* **84**, 2995–3007 (1979).
- Kieffer, H. H., Chase, S. C., Martin, T. Z., Miner, E. D. & Palluconi, F. D. *Science* **194**, 1341–1344 (1976).
- Anderson, D. M. & Tice, A. R. *J. molec. Evol.* **14**, 33–38 (1979).
- McCord, T. B., Huguenin, R. L. & Johnson, G. *Icarus* **31**, 293–314 (1977).
- Farmer, C. B., Davies, D. W., Holland, A. L., LaPorte, D. D. & Doms, P. E. *J. geophys. Res.* **82**, 4255 (1977).
- Hess, S. L., Leovy, C. B., Ryan, J. A. & Tillman, J. E. *J. geophys. Res.* **82**, 4559–4574 (1977).
- McGinnis, L. O. *Proc. 2nd Colloq. Plan. Water Hanover*, New Hampshire, 7–10 (1978).
- Brass, G. W. *Icarus* **42**, 20–28 (1980).

MATTERS ARISING

Superoxide involvement in negative air ion effects

INTEREST in research on the biological effects of small negative air ions has been due to the variety of claims made for them, ranging from lethal effects on micro-organisms to therapeutic effects on humans¹⁻³. In spite of the promising results reported, there has been a lack of understanding of the underlying biochemical basis. This is not surprising because the exact nature of the air ions has not been defined by those concerned. Consequently, there has been no investigation of the interaction between well defined ionized air species and biological systems of different degrees of complexity, which could have provided the basis for understanding the possible biological impact. Recently, however, experiments have been reported by Kellogg *et al.*⁴ which "support the concept that the negative air ions responsible for bacterial death consist solely or in part of hydrated superoxide anions". This conclusion was based on the observation that the enzyme superoxide dismutase (SOD) completely protected bacteria against killing by air ionization. This enzyme is known to remove O_2^- efficiently by catalysing its disproportionation into hydrogen peroxide and molecular oxygen. The negative air ions in that study were produced by a corona discharge-type negative ion generator.

We suggest caution in adopting this conclusion. The negative ion species generated in air by corona discharge have been identified accurately by mass spectrometry⁵. At atmospheric pressure, in static or dynamic conditions, CO_3^- is the dominant species produced. Such ions were found to be due to small concentrations of CO_2 in air. The CO_2 apparently reacts with O^- or O_3^- to form CO_3^- . It was estimated that only about 40 p.p.m. of CO_2 is needed at atmospheric pressure to convert all oxygen ions to CO_3^- species. Superoxide anion, O_2^- , was detected in small amounts only at a gas pressure lower than 200 torr, a maximum of about 15% O_2^- being produced at 20 torr.

We know of no published report that CO_3^- can be converted to O_2^- in water. We have tested⁶ the possible generation of O_2^- from ionized air in aqueous medium by simple and straightforward chemical assay such as reduction of nitroblue tetrazolium (NBT) to the corresponding formazan⁷. A 5-day exposure of an aqueous solution of NBT (10^{-5} M) to the ion generator (Modulon Corona discharge-type operated at a maximum discharge current of 0.1 mA) provided no indication of conversion to formazan. Also, no evidence could be obtained of the production of hydrogen peroxide, the product of O_2^- self-disproportionation

($k = 4 \times 10^5 \text{ M}^{-1} \text{ s}^{-1}$, ref. 8) in water exposed to the ion generator (1 ml water, 10 cm distance from the needle, 24-h exposure).

Finally, the negative ion flux in the work of Kellogg *et al.*⁴ was $5 \mu\text{A}$. This corresponds to about 3×10^{13} electron charges per s or 1.1×10^{17} electron charges per h. Even if all the negative charges produced O_2^- , and neglecting its decay, this is equivalent to $1.8 \times 10^{-7} \text{ mol } O_2^-$ per h. This differs significantly from $3 \times 10^{-4} \text{ mol } O_2^-$ per h as the stated level of O_2^- in the same work.

We have no ready explanation for the protection by SOD against bacterial killing by negative air ions observed by Kellogg *et al.*⁴. Only by carefully duplicating their experiments and manipulating the variables in their system could one hope to explain the observed data. Whatever the explanation, however, it is very unlikely to be mediated by superoxide radical anion.

Regarding the general mechanism of action of negative ion generators (which are almost exclusively of corona discharge-type), we find it surprising that no attention has been paid to the striking resemblance between them and electrostatic precipitators. These latter devices, well known in the chemical industry, efficiently collect dust, fumes and mists by charging the particles in an electrostatic field created by negative corona discharge electrodes⁹.

We suggest that the same effects attributed to some undefined small negative air ions could be explained by the electrical field produced by the generation of ions. Such a field could cause, for example, the enhanced migration of microbial cells, reminiscent of electrophoresis, which ends in agglutination or adsorption to the walls, resulting in a lower plate count.

The claimed feeling of well-being produced by ion generators¹⁻³ could be due to an atmosphere free of aerosols and dirt, cleaned by electrostatic precipitation. Would not any other air-filtration device give the same results?

I. ROSENTHAL

*Division of Food Technology,
Agricultural Research Organization,
The Volcani Center, PO Box 6,
Bet Dagan, Israel*

and

*Department of Chemistry,
Nuclear Research Center-Negev,
PO Box 9001, Beer-Sheva, Israel*

E. BEN-HUR

*Department of Radiobiology,
Nuclear Research Center-Negev,
PO Box 9001, Beer-Sheva, Israel*

1. Knoll, M., Fichmeier, J. & Schön, R. W. *Adv. Electron. Electron Phys.* **19**, 178-244 (1964).

2. Gualtierotti, R., Kornbluh, I. H. & Sirtori, C. *Bioclimatology and Aeroionotherapy* (Carlo Elba Foundation, Milan, 1968).
3. Kreuger, A. P. & Reed, E. J. *Science* **193**, 1209-1213 (1976).
4. Kellogg, E. W., Yost, M. G., Barthakur, N. & Kreuger, A. P. *Nature* **281**, 400-401 (1979).
5. Shahin, M. M. *Appl. Opt. Suppl.* **3**, 106-110 (1969).
6. Rosenthal, I., Juven, B. J., Gordin, S. & Ben-Hur, E. J. *appl. Bact.* **46**, 451-454 (1979).
7. Beauchamp, C. O. & Fridovich, I. *Analyt. Biochem.* **44**, 276-287 (1971).
8. Bielski, B. H. J. & Allen, A. O. *J. phys. Chem.* **81**, 1048-1050 (1977).
9. Kirk, R. E. & Othmer, D. F. (eds) *Encyclopedia of Chemical Technology* Vol. 5, 646-665 (1954).

KELLOGG, YOST AND KRUEGER REPLY—We have several points to make in reply to the comments by Rosenthal and Ben-Hur. First, no one has yet definitely characterized the chemical nature of negative air ions. The work by Shahin¹ on the ionic species produced by corona discharge seems quite interesting, and he does clearly identify the predominant negative ion made in his system (at atmospheric pressure) as CO_3^- . However, his system differs from ours² in several ways, including the maximal current used (1 mA compared with our 0.01 mA) and the corona discharge voltage (12 kV compared with our 5 kV). We cannot determine whether these differences would affect the negative ions produced. However, even if our system did make CO_3^- as the principal negative ion species, Shahin's work does not preclude the production of HO_2 , which as an uncharged species would go undetected by mass spectroscopic methods, and which would give rise to O_2^- in solution.

We do know of a number of studies that support the $O_2^-(H_2O)_n$ negative air ion concept. For example, Mohnen³ summarized his work as follows: "An attempt has been made to determine theoretically a reaction scheme to explain the formation of negative small ions in the troposphere and lower stratosphere. The terminal ions are found to be hydrated O_2^- and CO_4^- . The degree of hydration depends on temperature, and the mean number of water molecules attached to those ions is of the order of 3 or 4. As soon as the water vapor concentration is less than the concentration of carbon dioxide (about 8 km), the terminal ion is found to be hydrated CO_4^- . The reaction scheme is based on measured and reasonably estimated reaction rate constants determined by using U.S. standard atmosphere for the neutral gas composition." Regardless of the identity of negative ions in air, this does not preclude their transformation to other ions, such as O_2^- on addition to solution. Our data, which showed strong protection against negative ion bacterial kill by superoxide dismutase (SOD), but not by denatured SOD, strongly indicated O_2^- involvement in this negative ion effect, which occurred whether one considers negative air ions as $O_2^-(H_2O)_n$ or not.

The negative air ion system used by Rosenthal *et al.*⁴ to test for O_2^- production, by looking for nitroblue tetrazolium (NBT) reduction, or H_2O_2 accumulation, had several gross inadequacies. They neither grounded, nor even stirred, their target solutions and could not in fact even demonstrate a bactericidal effect of negative air ions, an effect repeatedly confirmed in numerous published reports. Our own work has shown that the effect of negative ions on NBT gives results that seem neither simple nor straightforward. We found that although negative ion treatment did not cause reduction of NBT to the formazan, it did cause the NBT solution to change spectrally, while concomitantly losing its susceptibility to reduction to the formazan by dithionite. Although SOD completely prevented this effect, we also found that denatured SOD, and even bovine serum albumin, gave complete protection. Our results with NBT, although unexplained, offer no evidence either for or against superoxide involvement in negative air ion effects.

Similarly, the failure of Rosenthal *et al.* to find evidence for H_2O_2 production in solutions exposed to negative ions does not in any way argue against O_2^- involvement. They failed to understand that one cannot simply 'add' negative charges indefinitely to a target solution, and as their system consisted simply of H_2O exposed to an ion generator without grounding, or even stirring, one would not expect to find sufficient H_2O_2 accumulation for measurement by even the most sensitive methods, let alone less sensitive chemical methods such as the iodometric that they used.

We apologize that we did inadvertently give an incorrect value for the maximal possible flux of O_2^- in our system— $0.18 \mu M h^{-1}$ gives a correct estimation of this flux. Over a 5-h period, this gives a ratio of $\approx 1.4 \times 10^{10}$ negative ions per bacterium. Although the mechanism of bacterial kill remains obscure, it does not seem to occur by agglutination due to electrostatic effects, because early in our experimental series we observed no agglutination of samples in direct microscopic examination. Furthermore, previous studies using polonium-210 or tritium as ion sources have also demonstrated bacterial kill, despite having a field substantially lower than that seen in corona discharge generators⁵. Earlier work using corona discharge generators⁶ found a reversal of damage to cells on exposure to intense visible light, an effect which should not occur with agglutination. Finally, the time required for cell death was much less than the minimal time for agglutination predicted by the von Smoluchowski equation⁷. Many other studies have shown the varied biological activity of negative air ions⁸⁻¹³, so the suggestion by Rosenthal and Ben-Hur that all negative ion effects are derived from electric field or electrostatic pre-

cipitation effects appears at odds with the experimental facts of the literature.

E. W. KELLOGG III
Membrane Bioenergetics Group,
Lawrence Berkeley Laboratory
and
Department of Physiology-Anatomy,
University of California,
Berkeley, California 94720

M. G. YOST
A. P. KRUEGER
Department of Biomedical and
Environmental Health Sciences,
School of Public Health,
University of California,
Berkeley, California 94720

- Shahin, M. M. *Appl. Opt. Suppl.* **3**, 106-110 (1969).
- Kellogg, E. W. III, Yost, M. G., Barthakur, N. & Krueger, A. P. *Nature* **281**, 400-401 (1979).
- Mohnen, V. A. *Geophys. Res. J.* **75**, 1717-1721 (1970).
- Rosenthal, I., Juven, B. J., Gordin, S. & Ben-Hur, R. J. *appl. Bact.* **46**, 451-454 (1979).
- Krueger, A. P., Smith, R. F. & Go, I. G. *J. gen. Physiol.* **41**, 359-381 (1957).
- Krueger, A. P., Reed, E. J., Brook, K. B. & Day, M. B. *Int. J. Biometeor.* **19**, 65-71 (1975).
- von Smoluchowski, *Phys. Z.* **17**, 557, 583 (1916).
- Krueger, A. P., Strubbe, A. E., Yost, M. G. & Reed, E. J. *Int. J. Biometeor.* **22**, 202-212 (1978).
- Krueger, A. P., Kotaka, S. & Andriese, P. C. *Int. J. Biometeor.* **9**, 201-209 (1965).
- Kotaka, S., Krueger, A. P. & Andriese, P. C. *Int. J. Biometeor.* **12**, 85-92 (1968).
- Krueger, A. P., Andriese, P. C. & Kotaka, S. *Int. J. Biometeor.* **12**, 225-239 (1968).
- Gilbert, G. O. *Int. J. Biometeor.* **17**, 267-275 (1973).
- Frey, A. H. *J. comp. Physiol. Psychol.* **63**, 121-125 (1967).

Role of PGE₂ in anion exchange in gastric mucosa

It has been suggested by Schiessel *et al.*¹ that a prostaglandin (PGE₂) increases Cl^- - HCO_3^- exchange in gastric mucosa, thereby protecting surface cells against excessive back-diffusion of H^+ from the lumen. Their evidence is necessarily indirect, because it is not known whether such exchange occurs at the nutrient membrane of the surface cell (as demanded by their hypothesis) or, indeed, whether exchange diffusion is a characteristic of this cell type. The authors do show an increase in Cl^- fluxes across the isolated mucosa due to PGE₂, and because anion exchange is rather nonspecific in this preparation², as in red blood cells^{3,4}, it is reasonable to suppose that exchange of Cl^- for HCO_3^- is also increased.

The authors, however, are guilty of a logical *non sequitur*: they deduce that PGE₂ increases exchange of cell Cl^- for nutrient HCO_3^- from a negative finding, that PGE₂ does not affect the depression in nutrient to secretory Cl^- flux due to the removal of secretory Cl^- (trans-concentration effect⁵). Exchange of Cl^- across the epithelium requires a suitable partner anion, yielding a null electrical current. In the secretory solution, the authors use replacement anions (isethionate, sulphate) which do not exchange for Cl^- , and this fact cannot be altered by a drug such as PGE₂.

The criticism here raises a pertinent question: does a modest level of HCO_3^- in

the secretory solution reduce the trans-concentration effect, and does PGE₂ then further reduce the effect? Evidence on this question would bear more directly on the authors' hypothesis.

RICHARD P. DURBIN
Department of Anatomy, University of
Umeå, Umeå, Sweden

- Schiessel, R., Matthews, J., Barzilai, A., Merhav, A. & Silen, W. *Nature* **283**, 671-673 (1980).
- Durbin, R. P., Kitahara, S., Stahlmann, K. & Heinz, E. *Am. J. Physiol.* **207**, 1177-1180 (1964).
- Gunn, R. B., Dalmark, M., Tosteson, D. C. & Wieth, J. O. *J. gen. Physiol.* **61**, 185-206 (1973).
- Dalmark, M. *J. gen. Physiol.* **67**, 223-234 (1976).
- Heinz, E. & Durbin, R. P. *J. gen. Physiol.* **41**, 101-117 (1957).

SCHIESEL ET AL. REPLY—We acknowledge the validity of the argument that PGE₂ cannot alter the trans-concentration effect because replacement anions which do not exchange for Cl^- , were used in the secretory solution, and we were aware of that fact. In these experimental conditions, however, failure of PGE₂ to affect J_{Cl^-} (the flux of $^{36}Cl^-$ measured from nutrient to secretory solutions) more than about the 50% usually found in control experiments in our laboratory suggests that the PGE₂ does not affect either of the remaining components of Cl^- flux—the active or diffusional components. An unfortunate phrasing of the sentence describing our findings may have led to the misunderstanding that we believed these experiments rigorously excluded the possibility that PGE₂ affects exchange diffusion of Cl^- .

We agree that much of the evidence for our hypothesis is necessarily indirect. The finding that PGE₂ did not sustain Cl^- flux in metiamide-inhibited tissues when HCO_3^- was absent from the nutrient solution supports this hypothesis, however. It is also possible that PGE₂ causes an increase in Cl^- permeability without change in exchange diffusion of HCO_3^-/Cl^- exchange.

The suggested experiments would be interesting, but in our opinion difficult to interpret and not definitive in excluding an effect of PGE₂ on exchange diffusion of Cl^- . If all three components of the Cl^- flux are operative when a suitable exchangeable anion (not Cl^-) is in the secretory solution, stimulation of Cl^- flux by PGE₂ would not specifically define the effect of PGE₂ on any of the three components of Cl^- flux. Conversely, failure of PGE₂ to stimulate Cl^- flux in these conditions might be interpreted as evidence that either a direct Cl^- - Cl^- exchange at the secretory membrane is required for the PGE₂ effect or that the non- Cl^- anion inhibited the effect of PGE₂.

WILLIAM SILEN
RUDOLF SCHIESEL
JEFFREY MATTHEWS
AMI BARZILAI
ARYEH MERHAV

Departments of Surgery,
Harvard Medical School,
Beth Israel Hospital
Boston, Massachusetts 02215

BOOK REVIEWS

Tertiary sector of the science industry

Edward Shils

THE activities of thinking and observing which are involved in the pursuit of scientific knowledge occur in a setting of many other activities and institutions which are not themselves scientific knowledge but which are indispensable to it. These include the intellectual relationships among individual scientists, within institutions, within and between countries; the traditions of institutions — such as laboratories — in which scientific work is done; the organs and institutions through which scientists communicate the results of their research; and the processes through which problems are selected and knowledge is sifted, rejected, affirmed, extended, elaborated and revised. It is in these actions and places that the “public” character of science is exhibited. They constitute what has come to be known as “the scientific community”, a complex made up of as many scientific sub-communities as there are fields of specialization, and institutions and countries in which scientific work is done.

In addition, there are the institutions and arrangements which bring scientific traditions into the focus of attention of young persons who might become scientists, and those processes by which aspirant scientists are inducted into the intellectual and moral discipline of scientific work. Since scientific work requires apparatus and time free from other obligations, the institutional arrangements whereby scientists are enabled to live and to have what they need to do scientific work are likewise parts of this immediately necessary complement to the advancement of scientific knowledge.

Around all this there have grown up a large number of supplementary activities which are not integral or immediately necessary to the pursuit of scientific knowledge and its incorporation, by application, into technology. Among these are the historiography of science and medicine and of the technology used in economic life and warfare, systematic analysis of the logical and epistemological structure of scientific propositions, reflection on the moral and other merits and demerits of scientific knowledge and of technology in

A Guide to the Culture of Science, Technology, and Medicine. Edited by Paul T. Durbin. (Free Press/Collier Macmillan: 1980.) \$45, £30.

their many forms, the sociological study of the structure of scientific communities and of particular scientific institutions, and discussions about whether and how science should be governed by authorities outside itself. Some of this penumbral intellectual effort has been carried on by academics. Some of it has become established in academic institutions as parts of existing disciplines such as history, philosophy and sociology. Some is part of journalism and of amateur scholarship.

Finally, at some distance from immediate involvement with the pursuit of scientific knowledge, there are the ramifications of science into the way of life and outlook of the ordinary members of society — its impact on their moral notions and conduct, their cosmological and religious beliefs, their conceptions of mankind and the world, and their ideas about and knowledge of science itself.

All of this vast complex may be called “the culture of science”. It is an immense subject, deserving of studious attention because it permeates so much of modern civilization.

A Guide to the Culture of Science, Technology and Medicine is a collaborative attempt to summarize the state of knowledge regarding the culture of science. It is a thick volume of nine parts, ranging in length from 30 to about 60 pages. The parts deal with the history of science, written by Arnold Thackray; the history of technology, by Carroll Pursell; the history of medicine, by Gert Brieger; the philosophy of science in its “historical, social and value aspects”, by Alex Michalos; the philosophy of medicine, by H. Tristram Engelhardt and Edmund Erde; the sociology of science and technology, by Jerry Gaston; medical sociology and science and technology in medicine, by Linda Aiken and Howard Freeman; and science policy studies by Diane Crane. Each essay is accompanied by a long and excellent bibliography.

As is inevitable with such a work, the outcome is very uneven. Most noticeably, the quality of the contributions is unrelated to their length. Two of the least valuable — those on the history of technology and the

philosophy of science — are respectively the shortest and nearly the longest in the volume. The long essay on the philosophy of science contains less about the culture of science than any of the others; it is a series of mechanically assembled summaries and bibliographical references, some of them patently second-hand. Some, but not all of the defects of this particular essay are attributable to the author; they are also due to the dreary literature which he has to summarize, a literature which is self-contained, having little to do with what scientists do and know.

It is not accidental that the richest contributions are those by Brieger, by Engelhardt and Erde, and by Aiken and Freeman. Each of these essays benefits from the authors’ intimacy with medical practice and biomedical science, and from the long traditions of learning and reflection by physicians and biomedical scientists; in these there is an exceptional awareness of moral issues. Exceptional, because although the contributors to the volume were apparently enjoined by the editorial board to pay particular attention to moral issues and questions of “values”, most of the authors deal with them in a perfunctory way, as if they were acting under orders and did not have their heart in it. It is only in those essays dealing with medicine — and in Mitcham’s essay on the philosophy of technology — that the moral or “value” problems are treated as genuine issues which practitioners and investigators are under compelling obligation to reflect upon, and to take into account in their actions.

How are we to explain this difference between the humdrum and insightful character of some of these essays and the lively realism of others? The answer to this seems to me to lie in the fact that, for some of the authors, the “culture of science, technology and medicine” is something to be written about from the outside, as an academic quasi-discipline with its own self-justifying traditions. They regard it as something to be studied and written about, but not as something in which they participate. This bespeaks a deficiency in their understanding of the concept. They do not really see the literature of their subjects as parts of the “culture” — despite the title of the book to which they have contributed. The centrality of scientific knowledge to the “culture of

Edward Shils is Professor of Sociology and Social Thought at the University of Chicago and editor of *Minerva*. His most recent book is *The Calling of Sociology* (University of Chicago Press, 1980).

science" seems to have escaped some of the contributors to this book, not least the sociologists among them.

Here, a word should be said about the contributions of sociology to the present volume. In the essay which deals with medical sociology and science and technology in medicine, the authors devote less than one-fourth of their space to the former and, although they deal with it appreciatively, they see that it touches only spottily on their topic. They point out that it is guided by the tradition of sociology as an academic discipline. What they are indicating is that medical sociology, with all its merits, has not really become part of the culture of science, technology and medicine. The same must be said about the way in which the sociology of science and technology is presented here.

The editors themselves must bear some of the onus for this faulty understanding of the culture of science. Otherwise how could they have omitted any treatment of the impact of scientific knowledge on the

outlook and conduct of ordinary members of modern societies? There is nothing in the book about that. Nor is there anything about scientific education. And, except for passing comments in the essays on the history of science and medicine, there is no serious consideration of the process of the growth of scientific knowledge. The editors have found the right name for their subject but they have not found the subject to which to apply it.

In the preface it is stated that "more than anything else, the volume should be useful to teachers of interdisciplinary courses in technology and values or science and society". If their expectation is realized, it will help to maintain such courses of study in their present superficiality and in their arbitrary and distorted conception of their subject. This book has considerable value but it is not a guide for those who are too isolated for their own good from science, technology and medicine, and who do not understand their own peculiar place in its culture. □

presented in this review, unfortunately, is based on the authors' own work, with the consequence that the chapter is too short to cover these interesting viruses in sufficient depth.

The final contribution is a review on the prospects in human retrovirus research. It does, however, present a somewhat unusual perspective as it is more cautious than previous reviews on the subject. The chapter is well balanced and covers the appropriate material comprehensively.

Overall, the book is an adequate compilation of many useful reviews. What is missing, however, is a final section that brings together the diverse ideas and information — instead of leaving the chapters as separate entities — and a discussion of the direction that future work in this field should take. It is also rather unfortunate that, at the time of completion of the book (early 1979), information on molecular cloning studies and structures of proviruses was still scanty, for additional material on these subjects would improve the book. This kind of drawback is perhaps inevitable in such multi-author reviews covering fast-moving areas of research. Nevertheless, this book should stand as a good source of references. It should also be of value to biochemists with a special interest in retroviruses and to pre- and postdoctoral fellows entering the field. □

Retrovirus research

Tak W. Mak

Molecular Biology of RNA Tumor Viruses. Edited by John R. Stephenson. Pp. 528. (Academic: 1980.) \$49.50, £34.20.

THE momentum that was given to retrovirus research by the discovery of reverse transcriptase is still strong. Today, ten years after that discovery, the field remains one of the most active areas in biology. The book, a collection of 12 reviews by as many groups, is an attempt to take stock and evaluate the progress in the field. In general, the reviews are informative though a few chapters are somewhat narrow in that they tend to dwell on the authors' own work. A more important criticism, however, is that the choice of topics suffers from lack of an overall balance that would accurately reflect the efforts in different aspects of retrovirus research. For example, the reviews tend to emphasize mammalian retroviruses, with a strong accent on murine type-C viruses, while omitting much of the avian viruses and their genetics.

The book begins with a good account of the historical development of retrovirus research, from the discovery of Rous sarcoma virus to the present. The three chapters which follow all deal with mammalian type-C retrovirus genes and their transmission. The first is a summary of the efforts to delineate further evolutionary relationships among primates and mice through studies of their endogenous retrovirus sequences. The next, on murine endogenous type-C viruses, is a comprehensive account of the

genetics and regulation of murine type-C retroviruses. It contains many helpful tables to aid readers in understanding this complicated field. The third of these opening contributions is on the integration and transmission of exogenous murine type-C virus. It contains some interesting information on germ line integration of exogenous virus. However, it seems rather brief and repetitive after the two preceding, rather ampler, chapters.

The next two contributions are on transforming viruses and retrovirus genomes, respectively. They are both useful reviews. They complement each other well and are presented at an appropriate level for such a volume. That which deals with retrovirus genomes, however, is too short to cover the genome structures of both avian and murine retroviruses. Expansion of this chapter or addition of another article to include more on avian retroviruses and the biochemistry of retroviruses would have been desirable.

The seventh and eighth chapters are on type-C viral proteins and their primary structures. They contain an in-depth look at the proteins of type-C retrovirus and would be useful for investigators interested in the subject. Next come two contributions on reverse transcriptase and heteroduplex mapping — both of which are informative and contain detailed surveys of reverse transcriptase enzymology and heteroduplex mapping of viral genomes. The eleventh chapter is a discussion of type-B and type-D retroviruses. Most of the information

Tak W. Mak is a Senior Staff Scientist at The Ontario Cancer Institute and an Associate Professor of Medical Biophysics at the University of Toronto.

Evolving theories

Adrian Friday

Phylogenetic Patterns and the Evolutionary Process. Method and Theory in Comparative Biology. By Niles Eldredge and Joel Cracraft. Pp. 349. (Columbia University Press: 1980.) \$27.50, £15.50.

ELDRIDGE and Cracraft's central theme is that estimates of the pattern of evolutionary history logically precede the making of hypotheses about the process of evolution.

Estimation of the pattern of evolutionary relationships is inevitably carried out by the comparison of similarities between organisms. To Eldredge and Cracraft, shared similarities do not all have equal status in this enterprise but, following the ideas of the late Willi Hennig, a way of organizing such information is to construct a series of nested sets each characterized by at least one evolutionary innovation (Hennig's "synapomorphy"). That these are innovations emphasizes to the authors the importance in such analysis of interpreting the pathway of change in biological characters. Decisions about the sequence of change can often be difficult

to make, but the stark alternative interpretation of the pattern of differences between organisms as a static, instantaneous phenomenon is clearly not one likely to appeal to evolutionary biologists. To quote Eldredge and Cracraft (p.55): "The concept of nested sets of synapomorphy makes sense only if it is assumed that change has taken place . . .".

The so-called "problem of homology" is presented as not being a problem when phylogenetic homologies are equated with the evolutionary innovations defining sets. Accordingly, evolutionary convergence is interpreted as non-homologous similarity: the sceptical might regard this as more a definition than a solution.

For Eldredge and Cracraft, a diagram summarizing set membership is a "cladogram" and includes no designation of ancestral versus descendant organisms. If such additional conjectures are made, the diagram then becomes a phylogenetic tree. The difference is seen as an important one: the cladogram being the more general statement. However, no mention is made of the valuable work of G.F. Estabrook which defines these concepts more precisely from a rather different viewpoint.

The authors present the structure of a cladogram as subject to rejection and replacement by an alternative hypothesis. In this way hypotheses of set membership can be corroborated. Rejection is achieved by the examination of more characters. Where these contradict the original partitioning, they may do so in concerted fashion and force a re-examination, not only of the old scheme of hierarchies, but also of our biological understanding of the characters used to define them. A statement that: "The criterion of parsimony specifies our acceptance of the least rejected hypothesis" (p.70) may seem unexceptionable to many, but: "In evolutionary terms, therefore, we prefer the cladistic hypothesis which minimizes convergence" (also p.70) appears unfortunate in view of the authors' otherwise clinical view of parsimony as an inevitable component of the hypothetico-deductive method rather than a principle of the evolutionary process.

In the final chapter the separate nature of microevolution, speciation and macroevolution is emphasized (with due acknowledgement to the ideas of Sewall Wright and S. N. Salthe). On this subject the book becomes repetitious and risks losing its momentum in nested reappraisals of previous sections. Certainly, a reading of such recent books as R. C. Lewontin's *The Genetic Basis of Evolutionary Change* (Columbia University Press, 1974), M. J. D. White's *Modes of Speciation* (Freeman, 1978) and S. M. Stanley's *Macroevolution* (Freeman, 1979) reinforces the lack of a unified theory underlying the events of evolution.

Ironically, the conflict between pattern and process can be brought out by con-

sidering the approach of population geneticists to the estimation of kinship relationships using gene frequency data measured in human populations, clearly a "microevolutionary" study (an example is given in E. A. Thompson's *Human Evolutionary Trees*; Cambridge University Press, 1975). This type of estimation is a statistical exercise involving the prior construction of a probabilistic model of change. When we attempt to apply such a methodologically rigorous approach to anatomical characters it is difficult to frame a plausible model for the process.

What Eldredge and Cracraft's book

suggests to me, in defiance of their declared theme, is that our lack of a model for the relevant processes of evolution hampers the convincing reconstruction of the pattern of evolutionary relationships. The majority of readers would probably not agree. In any case, the book is strongly recommended to allies and opponents alike. It solicits underlinings and exclamation marks and is mostly elegantly argued and enjoyable.

Adrian Friday is Assistant Curator of Vertebrates at the University Museum of Zoology, Cambridge, UK.

Humphry Davy and geological conundrums

G.Y. Craig

Humphry Davy on Geology: The 1805 Lectures for the General Audience. Edited by R. Siegfried and R.H. Dott. Pp.169. (University of Wisconsin Press: 1980.) \$17.50, £9.

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Sir Humphry Davy

HUMPHRY Davy, chemist and inventor, lectured on a variety of scientific subjects at the Royal Institution in London between 1801 and 1812. From 1805 onwards he gave a course of ten lectures in geology. Happily, full hand-written notes on eight of his lectures have been preserved, and with two further transcripts by a copyist now provide the only complete lecture course known among Davy's surviving papers.

In their introductory chapter the editors outline Davy's career and comment on his geological philosophy. Davy himself explains his stance in the first lecture:

Geology, perhaps more than any other department of natural philosophy, is a science of contemplation. It requires no experience or complicated apparatus, no minute processes upon the unknown properties of matter. It demands only an inquiring mind and senses alive to the facts

almost everywhere presented in nature. As it may be acquired without much difficulty, so it may be improved without much painful exertion.

Almost a century later Sir Archibald Geikie commented in his biography of Murchison that it also required a stout pair of legs!

Davy's style is clear, his writing elegant, his criticisms courteous, and throughout he reveals the caution typical of the chemist. The first four lectures, excluding a general introductory lecture, are largely concerned with a review of earlier theoretical speculations, ranging from Greek philosophers to late-eighteenth century writers. In the following six lectures he describes rocks and outcrops, both from his own observations and from those of the other naturalists. Conundrums are many, resolutions are few. Why should limestone stick more firmly to sandstone than to schist? There is a report of fishes in basalt! Is coal formed by heat from vegetable matter?—yet the strata above and below don't appear to have been subjected to heat. Sensible personal observations blended with second-hand reports, contradictory theories and a belief in a purposeful progression of life inevitably lead to difficulty. Davy does not attempt to develop an independent theory of his own, but with the critical and perhaps slightly superior view of an experimental chemist has as much to be wary about. Since he is not prepared to be a Neptunist, and is unwilling to be a Plutonist, he finds himself on occasions in an impossible position:

All the facts seem to render it probable that the basaltic rocks usually found in alternation with the common secondary rocks cannot be of igneous origin. But the proposition of the impossibility of the formation of this substance by fire must not be advanced.

Davy was not the first lecturer in geology in Britain; he was preceded by Walker at Edinburgh and by Beddoes at Oxford and Bristol. And although there was no

Mary Evans

text books from MACMILLAN PRESS

FROM DNA TO PROTEIN: The Transfer of Genetic Information

MARIA SZEKELY

This is the first textbook which discusses gene structure on the basis of recent discoveries of overlapping and split genes and which genetic information can be encoded into the genetic material and how it can be transferred to other materials.

A perfectly up-to-date account of our present knowledge of the structure and expression of genes in prokaryotic and eukaryotic organisms and the only available text showing the correlation between the structure and function of nucleic acids.

March, £18.00, 0 333 21836 1
£9.95, 0 333 21828 8

HUMAN REPRODUCTION AND DEVELOPMENTAL BIOLOGY

D. J. BEGLEY, J. A. FIRTH and
J. R. S. HOULT

An integrated approach to describe the cellular and physiological processes of human reproduction and embryological development in relation to their underlying mechanisms and to their practical applications in medicine. The book combines anatomy, biochemistry, physiology and pharmacology, making it more suitable for modern preclinical and human biology courses.

February, £20.00, 0 333 23423 5
£9.95, 0 333 23424 3

GYNAECOLOGICAL CYTOLOGY

MATHILDE E. BOON and
METTE LISE TABBERS-
BOUMEESTER

Dr Boon and Dr Tabbers-Boumeester have, over the past ten years, developed the course in diagnostic cytology which was first given in Leiden in 1969.

In this volume they present their insight into both the study of diagnostic cytology, and into the teaching of this increasingly important subject. To illustrate the modern concepts of cytopathology, examples of more advanced approaches are included: as well as scanning and electron microscopy there are details of DNA measurements, morphometry and immunopathology.

June, £25.00, 0 333 26117 8
£14.00, 0 333 26118 6

Macmillan Press books are available from leading booksellers. If you have difficulty obtaining these titles please write to Tony Watson, Macmillan Retail Ltd, Houndmills, Basingstoke, Hants, enclosing purchase price + 70p (p&p).

M MACMILLAN
PRESS

elementary textbook as we know it at the time, Playfair's *Illustrations of the Huttonian Theory* was certainly available. It would be interesting to know the names of the people who attended Davy's classes and the influence that he had on them. We know for instance from Geikie's biography that R.I. Murchison attended Davy's class in 1812 and that in 1823 Davy encouraged Murchison to go up to London and take some lectures in chemistry, etc. Davy's further irresistible bait was that he would soon get Murchison into the Royal Society (Davy was at that time its President!). Perhaps then Davy's greatest contribution to geology was to persuade Murchison to

abandon fox-hunting and take up geology instead.

Geologists and historians will be grateful to the editors for producing this book and continuing the tradition of academics in the mid-western United States in providing, indeed specializing in, transcripts and facsimile copies of the works of the late-eighteenth and early-nineteenth century philosophers of the Earth. Future historians of the history of science will no doubt one day wish to consider the reason for this phenomenon.

G. Y. Craig is Professor of Geology at the University of Edinburgh.

Cellular regulatory mechanisms

John Mowbray

Principles of Metabolic Control in Mammalian Systems. Edited by R.H. Herman, R.M. Cohn and P.D. McNamara. Pp.670. (Plenum: 1980.) \$35, £22.05.

THE editors of this large book are also the main contributors as authors or co-authors of 11 of the 15 chapters. Their aim, according to the preface, is to discern basic regulatory principles and to embody these in a theoretical approach which can be applied to any set of metabolic reactions. Anyone conversant with the development of metabolic regulation in recent years could be forgiven the thought that such an attempt is premature; close reading of the resulting volume has unfortunately done nothing to dispel my prejudice in this respect. This does not mean that the work should be dismissed out-of-hand, for there is a patent need to collate and review critically the diffuse literature on cellular regulation so that the various arguments may be widely appreciated and tested.

The subject is not one for those without a basic grounding in biochemistry, and yet many of the chapters are padded out with rather large amounts of elementary textbook information. The second chapter, for example, discusses metabolic networks from a non-linear, non-equilibrium, thermodynamic point of view in the first part and defines, as in a primer, hydrophobic, electrostatic, covalent and van der Waals interactions in the second part. In similar vein the special purpose of the book receives little or no contribution from the chapters on enzymes and coenzymes, protein synthesis, DNA replication and the cell cycle, membrane enzymes and membrane structure. All of these topics are better presented (there is a noticeable dearth of diagrams in this volume) in most textbooks for the beginner, and are readily accessible in expert reviews or monographs for the more

advanced reader. On the other hand, a topic not well covered as yet, protein degradation, is presented in a straightforward account with the methodological reservations adequately expounded.

Of the chapters devoted to regulatory mechanisms, the opening chapter in the book develops the hierarchical classification of metabolic processes from which the authors believe important principles can be deduced. The lists and tables are informative, albeit less than comprehensive. In a later contribution, modulation of enzyme activity is discussed without achieving the penetration of specialist monographs like those by Newsholme and Start (Wiley, 1973), or Ferdinand (Wiley, 1976), and it lacks the breadth of some recent reviews. The account of feedback/forward on complex pathways such as glycolysis strikes me as difficult to follow without the aid of diagrams unless one is already familiar with the examples. The chapter "Secretion Mechanisms" is a good, clear account with sufficient references for further reading, and that on the role of compartmentation is well written and nicely balanced with examples taken from key processes. Finally, the chapter dealing with hormone mechanisms has an interesting section on comparative biochemical endocrinology which is perhaps too complex for the novice, and a comprehensive though somewhat uncritical treatment of fast-acting hormones. By contrast the treatment of steroid hormones is cursory in the extreme, spoiling what is otherwise one of the better chapters.

In the end, while there are selected chapters one can recommend to one's students, this is a book for the specialist who can whin the grains of very real effort the editors have sown from the chaff of their self-indulgence.

John Mowbray is a Lecturer in Biochemistry at University College London.